SUN: Social and hUman ceNtered XR

-

A Horizon Europe Project Paving the Way for the Widespread Adoption of Extended and Virtual Worlds

Claudio Vairo, Giuseppe Caracciolo, Daniela Giorgi,
Daniele Leonardis, Lucia Vadicamo (Eds.)

edizioni
Consiglio Nazionale delle Ricerche

SUN: Social and hUman ceNtered XR

A Horizon Europe Project Paving the Way for the Widespread Adoption of Extended and Virtual Worlds

Edited by Claudio Vairo, Giuseppe Caracciolo, Daniela Giorgi,
Daniele Leonardis, and Lucia Vadicamo

Funded by
the European Union

Consiglio Nazionale delle Ricerche
Istituto di scienza e tecnologie dell'informazione

Series name
ISBN (electronic edition) 978-88-8080-801-5
DOI https://doi.org/10.32079/ISTI-BOOK-2025/001

www.edizioni.cnr.it

Layout and graphics by Lucia Vadicamo
Cover by Guido Dallago
Cover photo by Vincenzo Croce, featuring researcher Federica Serra from the Scuola Superiore Sant'Anna (SSSA) using one of the haptic devices developed within the SUN project.

This book has received financial support by the Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR - SUN project).
Views and opinions expressed in this book are those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the European Commission can be held responsible for them.

Published by

Piazzale Aldo Moro, 7
00185 Roma

www.edizioni.cnr.it
bookshop@cnr.it

Preface

Extended Reality (XR) is an emerging technology with promising potential in many fields, including healthcare, communication, and safety. However, overcoming XR's current limitations in providing authentic and interactive social environments is the key to fully unlock its potential. XR builds upon Augmented Reality (AR) and Mixed Reality (MR), which in turn builds on top of Virtual Reality (VR). VR allows users to access virtual worlds through headsets, smart devices, or computer screens, but it isolates them from the real environment. AR and MR, on the other hand, blend virtual and physical worlds, overlaying virtual elements into reality. In MR, users can interact with both physical and virtual objects, creating a more seamless experience. The term Extended Reality is often used as an umbrella concept for all the above, but it also goes further. XR includes perception within the virtual world, allowing for more natural and realistic interactions. Users can touch virtual objects, feel their weight, temperature, and texture, bringing the digital world closer to a real-life experience.

The main aim of the EU-funded SUN (Social and hUman ceNtered XR) project (https://www.sun-xr-project.eu) was to investigate and develop XR solutions that integrate the physical and the virtual world in a convincing and effective way, offering new opportunities for social and human interaction. SUN identified key limitations hindering the practical adoption of XR and addressed them through a multidisciplinary approach, where research and development were conducted synergistically across several complementary domains.

SUN explored and developed innovative approaches to *reconstructing, streaming, and enriching visual content* for XR. These advancements include artificial intelligence-based scalable methods for generating and improving 3D environments, objects, and avatars; high-performance solutions for interactive streaming in XR; and the application of artificial intelligence to analyze and enrich 3D assets. SUN investigated advanced solutions for *collaborating and interacting* in XR. The project introduced novel techniques for delivering thermal feedback to users, alongside innovative wearable haptic devices that enable more natural manipulation of virtual objects. Additionally, the project developed systems that support collaborative interaction among multiple users within a shared XR

session. SUN also explored and experimented with novel paradigms for *capturing user input and feedback*. Specifically, it proposed and demonstrated the use of electromyography to infer user movement intentions, alongside sensor and computer vision-based techniques for postural assessment and pose estimation. These efforts were complemented by solutions for hand gesture recognition and user emotion detection.

The components and solutions developed within SUN were integrated into the *SUN Integrated Platform*, an innovative technological framework designed to support the sustainable and secure development of XR applications. The platform incorporates solutions based on non-fungible tokens for managing XR assets, along with advanced techniques for detecting cyber threats in XR applications. SUN also addressed the *sustainability of XR solutions* by investigating viable business models and examining the legal, ethical, and societal dimensions of XR. This includes exploring the implications of processing personal and sensitive data collected through wearable devices, 3D acquisition technologies, and human-machine interaction systems.

Over the course of its three-year duration, the SUN project validated its technologies through real-world case studies focused on key application areas: XR for rehabilitation, XR for workplace safety and social interaction, and XR solutions designed to support individuals with mobility and verbal communication impairments. Validation activities were conducted at three distinct sites: the Rehabilitation Unit of Versilia Hospital in Lido di Camaiore (Italy), the shop floor of FACTOR in Valencia (Spain), and the Clinique Romande de Réadaptation of SUVA in Sion (Switzerland). In each case, the validation process actively involved real end-users, ensuring that the solutions were tested in authentic, practical contexts.

SUN has not only advanced the technological frontier of extended reality but has also demonstrated its tangible value in real-world contexts, paving the way for sustainable, ethical, and human-centered XR solutions. By integrating multidisciplinary research with practical validation, SUN leaves behind a robust foundation for future innovation, collaboration, and adoption of XR technologies across Europe and beyond.

In this book, we present the outcomes of the inspiring and ambitious journey undertaken by the SUN project. Our goal is to provide a comprehensive and coherent overview of the extensive research, development, and innovation carried out collaboratively by the project consortium. Through a multidisciplinary approach and close synergy among partners, SUN tackled complex challenges in XR, delivering groundbreaking solutions that reflect the dedication, creativity, and expertise of all the contributors.

November 2025

Giuseppe Amato
SUN Project Coordinator

Contents

SUN Technologies

Reconstructing, Streaming, and Enriching Visual Content 3

1 Improving the Creation of 3D assets 4

2 End2End Avatar Production Pipeline 19

3 Autonomous 3D Environment Exploration and Reconstruction 29

4 High-Performance Interactive Streaming 40

5 Open-Vocabulary Understanding of Objects and Scenes 49

Collaborating and Interacting 60

6 Thermal Feedback in Wearable Haptics 61

7 Wearable Haptics in Manipulation 71

8 Distributed Wearable Haptics 84

9 XR Collaboration and Gaze-Based Interaction 94

10 Task Optimization and Prioritization 108

Wearable and Vision-Based Monitoring Technologies for XR 120

11 EMG Decoding System for Hand and Wrist Kinematics 121

12 Postural Assessment and Monitoring of Body Kinematics 131

13 Multimodal Pose Estimation 142

14 Hand Gesture Recognition 151

15 Multimodal Emotion Recognition 161

SUN Platform and Cybersecurity 172

16 SUN Integrated Platform 173

17 Tokenized Platform for Customers Digital Assets Exchange 189

18 Cyber Threat Detection in XR 200

Human-Centered XR Scenarios and Real-World Case Studies

19 Humanity and Ethics Driven Scenarios 212

20 Extended Reality for Rehabilitation 230

21 Extended Reality for Safety and Social Interaction at Work 245

22 Extended Reality for People with Serious Mobility and Verbal Communication Diseases 253

Sustainability, Ethics, and Impact

23 Exploitation and Business Model 267

24 Impact: A Multi-Perspective View 285

25 Legal and Ethical Issues of SUN XR 301

SUN Technologies

SUN technologies enable a new generation of immersive, intelligent, and secure XR experiences. From realistic avatars and advanced 3D environments to seamless interaction and collaboration, these technologies drive the evolution of digital spaces. SUN's innovative tools for reconstructing, enriching, and streaming content make lifelike XR accessible and dynamic. By integrating natural collaboration tools—including haptic feedback, gaze tracking, and efficient task management—SUN promotes intuitive, shared XR experiences. Monitoring systems that leverage gesture, emotion, and movement recognition further enable adaptive, user-centered interactions. Underlying it all, the SUN Platform delivers robust cybersecurity and secure asset management designed for the evolving demands of XR, ensuring that these immersive environments are both open and trusted. Together, these advances open the door to digital worlds that feel real, collaborative, and secure.

Reconstructing, Streaming, and Enriching Visual Content

High-quality 3D assets are fundamental to creating immersive and interactive XR experiences, as they define the realism and usability of XR environments. This part introduces technologies for the generation and streaming of high-quality 3D content, encompassing different scales and semantics—from small objects to human avatars, up to large environments. SUN delivered innovative techniques for 3D asset creation, where both geometry and appearance are faithfully reconstructed. It defined techniques for reconstructing and rendering 3D scenes and navigating 3D environments based on semantic cues. Finally, SUN improved techniques for real-time, high-quality 3D content streaming. In addition, the project explored methods for scene understanding, including object detection and segmentation.

1. Improving the Creation of 3D assets

Daniela Giorgi[1], *Marco Callieri*[1], *Gianpaolo Palma*[1], *Massimiliano Corsini*[1], *Paolo Cignoni*[1], *Panagiotis Vrachnos*[2], *Spyridon Symeonidis*[2], *Sotiris Diplaris*[2], *and Stefanos Vrochidis*[2]

[1] Institute of Information Science and Technologies, National Research Council (CNR-ISTI), Italy
[2] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece

Abstract. The Social and hUman ceNtered XR (SUN) project promises Extended Reality (XR) solutions that integrate the physical with the virtual world. The quality and realism of the 3D digital models used to create immersive three-dimensional environments are pivotal to the quality of the experience, and improving this aspect can greatly enhance the user immersion and the effectiveness of XR applications. This chapter outlines the different approaches explored in the SUN project to improve the quality of these 3D assets. As there are multiple ways and workflows used to create, process, and prepare the 3D models for their use in interactive XR applications, it is not sensible to propose a single solution able to fit all possible scenarios, object classes, and technologies employed. For this reason, the project addressed individual, recurring problems and covered different stages of the workflow: from the digitization step to 3D model preparation. At the same time, part of the effort has also been devoted to explore different representations of 3D objects, not based on triangulated meshes, but relying on modern neural technologies, to provide possible alternative representations able to cover different use cases and specific scenarios.

1.1 Introduction

One of the essential components of an Extended Reality (XR) application is the 3D environment, which is populated by *assets*: 3D digital models prepared for use in game engines and XR applications. The interaction between the user and the virtual/augmented world depends on the quality of these 3D assets. When realism is required, digitizing objects from reality is one way to create such 3D assets, but the process, although capable of generating realistic digital models, is far from perfect and often produces incomplete 3D models or visual artifacts in their appearance.

This chapter presents different Artificial Intelligence (AI) techniques to improve the visual quality of 3D assets digitized from real-world objects. In a real-world XR development scenario, 3D assets are sometimes imported from external sources, and some other times created from scratch with a plethora of digitization methods and software tools. This makes it difficult to develop a single, all-purpose solution for improving these media. Moreover, different classes of objects might require different correction and improvement strategies. It made sense then to address specific problems in the creation and preparation of 3D assets and to cover different workflows.

The research focused on different phases of the digitization process [Maggiordomo et al. 2023; Callieri et al. 2025], but also on the basic representations for managing 3D assets, so as to:

- *Improve the creation of new 3D assets by intervening in their digitization process*. AI methods can be used to correct problems in the input of widely-used photogrammetry digitization tools, resulting in complete and visually-pleasing 3D models (Section 1.2);
- *Improve existing 3D assets by correcting visual errors and integrating missing areas*. When using digitized 3D models, it is often necessary to correct visual errors in their color representation. AI-based inpainting effectively improves the quality of texture mapping (Section 1.3);
- *Provide alternative workflows for the creation of 3D assets*. The idea is to explore different 3D representations, not based on triangulated meshes, but relying on modern neural technologies, to provide alternative solutions in specific use cases and scenarios (Section 1.4).

1.2 AI-Driven Specular Removal for Photogrammetric Reconstruction

Photogrammetry is one of the most used methods for the digitization of real objects, due to its versatility and cost-effectiveness. The digitization process via photogrammetry involves two major phases: capturing a dense set of photographs with a camera at close range, and the actual reconstruction, which uses multiple, overlapping images taken from different viewpoints to reconstruct the position of 3D surface points. The quality and completeness of the 3D reconstruction strongly depend on the quality of the input photos, and there are various possible problems in the input photos that can result in visual artifacts, errors, or missing areas in the final 3D model. One of the most common problems is the presence of specular highlights.

The approach followed is to train a neural network to analyze the input photos, finding and/or correcting the local visual problems that affect the photogrammetric pipeline. The architecture consists of a U-Net [Ronneberger et al. 2015], capable of effectively capturing local and global information. The U-Net architecture features an encoder-decoder structure with skip connections that downsamples the input image and then upsamples the feature maps to generate the output.

The problematic areas can be either corrected or used to generate masks, which are supported by most photogrammetric software. The double option of being able to mask out or correct the issues makes the process versatile and controllable. Masking out the problematic areas preserves the scientificity of the process, as all the resulting 3D data come from measurements; correcting the input images is expected to lead to a more complete model, at the price of having some data hallucinated by AI. Therefore, a combination of the two approaches (i.e., use masked data for geometric reconstruction and corrected data for texture generation) might provide a good middle ground.

The architecture is trained to learn to remove specular highlights from an image. The training uses a synthetic dataset composed of renderings of realistic objects with and without specular reflections. The trained network is then integrated into a simple tool that loads a number of images and applies the highlight removal process to all of them. In a digitization scenario, the user will load the entire photogrammetric input dataset and process it, thus obtaining a new dataset with a reduced presence of specular highlights (Figure 1.1): this processed dataset will then be used with any photogrammetric software, resulting in a more accurate and detailed 3D model.

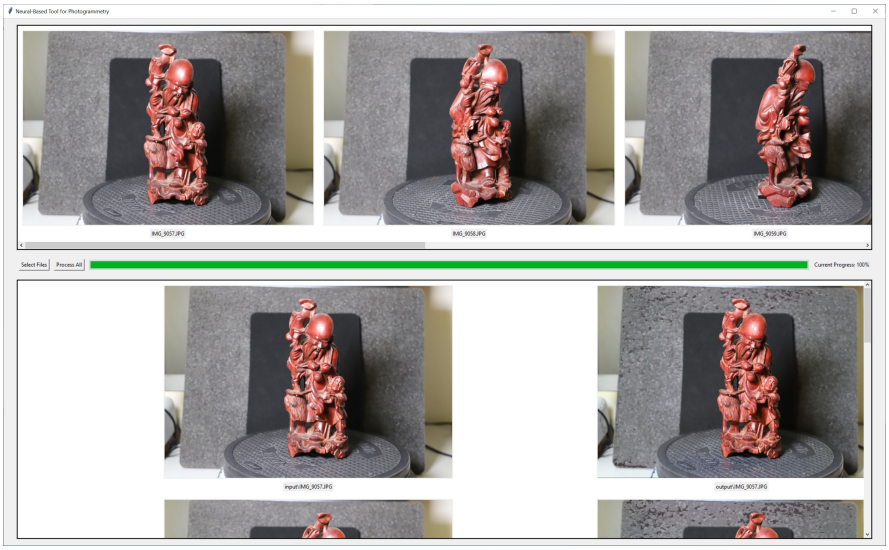


Figure 1.1: The AI image pre-processing tool used to prepare a photogrammetric dataset.

1.2.1 Dataset

We generate a synthetic image dataset using collected 3D models and a rendering engine. We chose Blender for its fast integrated rendering engine, capable of generating lifelike images with realistic illumination and material management, and the possibility to replicate arbitrary lighting/camera setup, and automate the rendering process with its Application Programming Interface (API).

The 3D models in the dataset must meet several criteria: they should be small objects suitable for photogrammetry, created through digitization, and rich in surface information. They should represent common photogrammetry targets, including everyday items, cultural heritage artifacts, and natural objects. They must feature well-defined geometries, realistic textures, and associated normal and specular maps to ensure accurate spatially-varying surface behavior. Additionally, they should exhibit a variety of materials, optical characteristics, shapes, colors, and details. Our dataset comprises 25 3D models meeting the specified criteria. Most of these models are sourced from SketchFab, an online platform for publishing 3D models; the models have a compatible license (free to use or Creative Commons) and are marked as usable for AI applications by the creator.

Each model is standardized in terms of size, orientation, position, and rendering materials. We use Python scripts and Blender's APIs to create scenes with different lighting setups. For each setup, Blender's Compositing feature allows the simultaneous

generation of the images needed to train our network: with highlights (diffuse + specular – input images), without highlights (diffuse only – target images), and highlights only (specular only – employed in the custom training loss defined in the following).

We use three lighting setups: a single spotlight, four area lights positioned far above the object, and three area lights placed closer around the object. These varying lighting setups are designed to enhance the network's generalization, as rendered images with diverse lighting conditions help the network learn to remove highlights under different scenarios. For camera placement, we use a hemispherical arrangement, with the camera moving around the object at various heights, pointing towards the object's center. This strategy is the most common in photogrammetric reconstruction. We set the number of cameras to 40 to ensure accurate 3D reconstruction and to facilitate learning of varied specular highlights. The rendering resolution is set to 2000×2000, sufficient to preserve details in the final images and maintain specular highlights of suitable size and complexity, while ensuring reasonable rendering and training times.

1.2.2 Custom loss

To fully exploit the training dataset, we design a custom loss made of two terms:

$$\mathcal{L} = \mathrm{MSE}(I, T) + \frac{k}{w} \sum_{i,j} \sigma(M_{i,j})(I_{i,j} - T_{i,j})^2 \tag{1.1}$$

where MSE is the Mean Squared Error value, M is the specular-only image, σ is a sigmoid activation function with high steepness, k is a constant (set equal to 10 in our experiments) and $w = \sum_{i,j} \sigma(M_{i,j})$. Specular highlights images serve as input images I, and diffuse component images as targets T.
The first term in the loss encourages the fidelity of output images to input ones, preventing hallucinations and preserving details. The second term penalizes differences between input and target images in correspondence with highlights.

1.2.3 Results

From a quantitative point of view, we evaluate the Structural Similarity Index Measure (SSIM) between the output images and the ground-truth images. SSIM assesses image similarity based on luminance, contrast, and structure. In the test set, SSIM is equal to 0.945 ± 0.08, indicating detail preservation and high fidelity of the corrected images to the input images. From a qualitative standpoint, we discuss results on real-world objects to assess the generalization ability. The objects have a variety of geometries, colors, or materials. Evaluation criteria include the extent of specular reflection removal, texture preservation, and overall visual fidelity.

Figure 1.2: Results of the Specular Removal on three synthetic objects (unseen in training). For each row: input image, our method, ground truth.

Figure 1.2 shows the results on three synthetic objects, not used for the training. Our customized U-Net is effective in removing specular reflections and preserving texture details, resulting in higher overall visual fidelity. The reference article [Callieri et al. 2025] presents results for real-world and synthetic objects, providing a comparison with state-of-the-art methods JSHDR, M2-Net, and DHAN-SHR. In all cases, our method significantly reduces highlights, making the processed images more suitable for subsequent stages of 3D reconstruction. This demonstrates the model effectiveness in handling real-world scenarios. In comparison, JSHDR generally produces lower quality results. M2-Net reduces the brightness of the specular, but does not correctly interpolate the image. DHAN-SHR produces results similar to our network, but more noisy in some cases, as shown by the head of the Buddha statue. Importantly, we are able to obtain comparative or even better results than DHAN-SHR with a much lighter archi-

tecture. Moreover, our performance is coherent across the entire set of input images: this is fundamental, since our goal is photogrammetric 3D reconstruction.

1.3 AI-based Texture Inpainting

The process of generating color information of digitized 3D models relies on projecting photos onto the reconstructed geometry and encoding the projected data into a texture. This process may generate errors in the final model in different ways: small inconsistencies in the camera position, uneven or changing lighting between photos, visual errors, or extraneous objects in the input photos. The result is local visual artifacts in the texture. AI-based inpainting is a common task that has been addressed using different approaches (e.g., encoder-decoder architecture, diffusion-based techniques), but it is necessary to find a way to apply this image-based tool in the fragmented, non-continuous, distorted parameterized space of a texture map.

Our solution is to apply inpainting to an ad hoc parametric domain that is defined on-the-fly for each inpainting operation. Given a region of the mesh that presents a texturing error, we create a local auxiliary parametrization of the affected faces and a region around them, and synthesize a texture for this region. This temporary texture is used to inpaint over the defective region, and the inpainted texels are transferred back into the original texture.

The auxiliary local parametrization is built by firstly creating a small patch starting from the problematic area using a local-global As Rigid As Possible (ARAP) parametrization [Liu et al. 2008], and then iteratively extending it, one triangle at a time, until a sufficiently large region around the target area is fully covered [Myles and Zorin 2012]. To perform the inpainting, we employ a close adaptation of a recently proposed Deep Convolutional inpainting network [Liu et al. 2018a], although any other method can be used, thanks to the modularity of the approach. This model uses an encoder-decoder architecture with skip-connections analogously to UNet [Ronneberger et al. 2015], and partial convolutions [Liu et al. 2018b] to improve performance with masked image inputs. The network takes as input the generated temporary texture and a binary mask denoting valid and invalid pixels, and outputs a reconstructed RGB image. The inpainting network was trained using a task-specific dataset, that is, texture patches created from scanned objects. For the training, we generate a collection of around 50000 patches sampled from a public benchmark of textured 3D real-world models [Maggiordomo et al. 2020].

Following these principles, we develop a texture repairing tool that locally unwraps the problematic areas, applies the inpainting, and integrates the result back in the

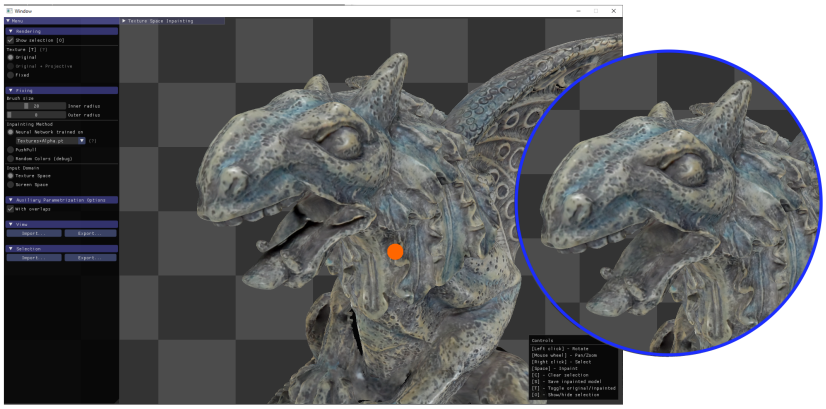


Figure 1.3: The AI inpainting tool correcting the texture map of a 3D-scanned Gargoyle statue.

Figure 1.4: Details of the texture correction on different kinds of 3D models.

texture. The tool works both automatically, detecting and correcting imperfections in the texture, and as an interactive tool with a paint/select interface to indicate which parts of the model need repairing, as shown in Figure 1.3. Results of the AI inpainting process can be seen in Figure 1.4.

1.4 Exploring Alternative Representations

Creating realistic 3D representations of real-world scenes remains a challenge due to the intricate demands of capturing high-detail appearance and geometry. Neural Radiance Fields (NeRFs) [Mildenhall et al. 2021] have recently emerged as a revolutionary approach to 3D reconstruction, presenting a distinct learning-based method for capturing and synthesizing complex 3D scenes. Traditional methods that rely on discrete voxels or point clouds often struggle to accurately reconstruct complex scenes and synthesize novel views. In contrast, NeRFs operate by modeling the volumetric scene as a continuous field of radiance, allowing for unparalleled accuracy and realism. The core idea of NeRF relies on representing the radiance field of a 3D scene, implicitly,

through a multi-layer perceptron (MLP), which enables the generation of high-quality images from novel viewpoints. Unlike NeRF's implicit representation, the 3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] method offers explicit scene representation and novel view synthesis, without the necessity of a neural network. This prominent technique utilizes a set of Gaussian ellipsoids and rasterizes them into images, allowing for efficient, real-time rendering while accurately representing both geometry and appearance with photorealism. Recently, the integration of super-resolution techniques with neural rendering approaches like NeRF and 3DGS has been explored to enhance the quality of novel view synthesis from low-resolution inputs. These enhancements aim to overcome common limitations, such as blurry or aliased outputs when rendering at resolutions higher than the input. In the context of the SUN project, such advancements were considered to represent 3D-related scenes and also to improve the fidelity and applicability of neural rendering methods in real-world scenarios.

1.4.1 Neural Representations

The main approach of NeRF leverages an MLP to encode both the scene geometry and appearance as a volumetric function, given a collection of 2D images and their corresponding camera poses. The network learns to map directly from viewing direction and spatial location to opacity and color, using volume rendering. While NeRF achieves exceptional rendering quality by sampling points along rays in 3D space, it requires higher training and rendering time due to the computational complexity of its volumetric ray-casting. On the other hand, 3D Gaussian Splatting directly optimizes Gaussian ellipsoids and projects them onto pixels using a rasterization-based rendering approach, which eliminates the need for dense sampling. With this rasterization technique, it not only achieves high-quality novel view synthesis but also significantly accelerates the training process, offering real-time rendering. Moreover, 3DGS models opacity values directly (unlike NeRF, which derives opacity from transformed density values), further optimizing the process while maintaining high visual fidelity.

A prerequisite for both NeRF and 3DGS methods is the accurate estimation of camera parameters and localization. This estimation is achieved through Structure from Motion (SfM) techniques, where camera positions are represented in 3D space. For this purpose, two different methods were tested, named COLMAP [Schonberger and Frahm 2016] and Hierarchical Localization [Sarlin et al. 2019].

In the context of the project Pilot 2 on safety and social Interaction at Work (Chapter 21), experiments were carried out utilizing the Nerfacto model [Tancik et al. 2023] and its extension, Nerfacto-huge (see Figure 1.5). Initially, an investigation was conducted to acquire web data illustrating scenes related to shop floor video captures.

(a) (b)

Figure 1.5: Experimental results of NeRF in data provided by FACTOR. (a) nerfacto model result on the metallic item sample. (b) nerfacto-huge result on the helmet sample

Due to limitations in web data quality and viewing angles, the project partners captured and delivered customized recordings with various objects relevant to the scenario (e.g., a helmet).

In the context of the project Pilot 3 on the rehabilitation of people with mobility and verbal communication diseases (Chapter 22), both NeRF and 3DGS methods were utilized on custom captured video via smartphone, from which an optimal number of frames was extracted based on the complexity and content of the scene. Figure 1.6 presents a qualitative evaluation of the radiance-based reconstruction techniques implemented, namely, a visual comparison between the ground truth image, the Nerfacto output, and the Splatfacto output. The Nerfacto model output contains more noise, while the Splatfacto model produced more detailed and compact results, particularly in the region around the cup.

To further improve the quality of novel view synthesis, particularly from low-resolution inputs, super-resolution methods were investigated alongside NeRF and 3DGS. Two advanced techniques were explored: NeRF Super-Resolution and Super-Resolution 3D Gaussian Splatting.

1.4.2 NeRF Super-Resolution

NeRF Super-Resolution (NeRF-SR) [Wang et al. 2022] is a supersampling technique that enforces sub-pixel level multi-view consistency and introduces a refinement network based on depth maps to transfer high-frequency information from a single high-resolution reference image. This approach enhances output resolution without requiring extensive high-resolution training data.

(a) GT image **(b)** Nerfacto output **(c)** Splatfacto output

Figure 1.6: Visual comparison between ground truth image, Nerfacto output, and Splatfacto output. Please zoom in for a better inspection of the results.

1.4.3 Super-Resolution 3D Gaussian Splatting

Super-Resolution 3D Gaussian Splatting [Feng et al. 2024] (SRGS) approach extends 3DGS to high-resolution novel view synthesis (HRNVS) from low-resolution input by combining two key mechanisms:

- *Super-resolution Gaussian Densification*, which encourages splitting and densification of primitives with sub-pixel constraints;
- *Texture-Guided Gaussian Learning*, which leverages an external pre-trained 2D super-resolution model to recover high-resolution textures from low-resolution input views.

This joint optimization effectively bridges the gap between low-resolution inputs and high-resolution outputs. Figure 1.7 presents a qualitative comparison between the NeRF-SR method and the original NeRF implementation (Vanilla NeRF) on a custom-captured video.

A corresponding quantitative evaluation in terms of Peak Signal-to-Noise Ratio (PSNR) is provided in Table 1.1. While the visual results show no significant differences, the quantitative analysis indicates that NeRF-SR yields improved high-resolution output quality when operating on low-resolution inputs.

(a) Vanilla NeRF: Input/Output: 270×480 **(b)** NeRF-SR: Input/Output: 135×240 **(c)** NeRF-SR: Input 135×240. Output: 270×480

Figure 1.7: Visual comparison between original implementation Vanilla NeRF, the output of NeRF-SR model at the same resolution as the input, and the output of NeRF-SR model at $2\times$ the input resolution.

Table 1.1: Quantitative comparison between Vanilla NeRF and NeRF-SR

Model	PSNR
Vanilla NeRF Fine	36.7524
NeRF-SR Fine	36.9469

To evaluate the effectiveness of the SRGS method, a performance comparison was conducted against the baseline 3DGS on the same data. In both experiments, Gaussian primitives were trained using low-resolution views obtained by downsampling the original images by a factor of $8\times$ (to 270×480). Subsequently, high-resolution rendering was performed at 1080×1920, corresponding to a $4\times$ upscaling of the input resolution. Figure 1.8 and Table 1.2 demonstrate qualitative and quantitative comparison results, respectively. The results highlight SRGS performance surpassing the 3DGS method in high-resolution rendering in the metrics of PSNR, Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS).

(a) Ground truth image **(b)** 3DGS high-res image **(c)** SRGS high-res image

Figure 1.8: Visual comparison between ground truth image, the output of baseline 3DGS method, and the output of SRGS method rendering at $4\times$ the input resolution.

Table 1.2: Quantitative metrics for high-resolution novel view synthesis

Model	SSIM	PSNR	LPIPS
3DGS	0.9264	28.4266	0.2703
SRGS	0.9458 ↑	30.9879 ↑	0.2499 ↓

1.5 Conclusions

The work carried out in the SUN Project addressed one of the bottlenecks of the creation of XR applications: the production of 3D assets for setting up a realistic immersive environment. The results of the proposed techniques show improvements in the visual quality of 3D models, resulting in more realistic assets. This is especially important in those XR applications that rely on the perception of realism to establish a specific relationship between the user and the environment. Working on specific issues and asset-producing workflows proved to be the correct strategy, as in this way, it is possible to provide a more comprehensive and modular contribution to the process.

While these activities were successful in improving the quality, consistency, and even the representation methods of 3D assets, the needs of an immersive, XR tool

go beyond the visual layer. An asset, especially the ones that will be manipulated by the users, will need to be enriched with physical properties, starting from the basic information about weight and mass distribution. Being able to estimate such properties while digitizing the real object, thus resulting in a really integrated, all-round asset creation workflow, would increase the efficiency of the authoring process of XR applications. The SUN Project carried out experiments in this direction, but more work will be necessary to obtain usable results, ready to be integrated in immersive applications, also for the lack of standards to store these properties and import them inside the XR frameworks.

References

Callieri, Marco, Massimiliano Corsini, Somnath Dutta, Daniela Giorgi, and Marco Sorrenti (2025). "AI-Driven Specular Removal for 3D Asset Creation". In: *2025 25th International Conference on Digital Signal Processing (DSP)*, pp. 1–5.

Feng, Xiang, Yongbo He, Yubo Wang, Yan Yang, Wen Li, Yifei Chen, Zhenzhong Kuang, Jianping Fan, Yu Jun, et al. (2024). "SRGS: Super-Resolution 3D Gaussian Splatting". In: *arXiv preprint arXiv:2404.10318*. URL: https://arxiv.org/abs/2404.10318.

Kerbl, Bernhard, Georgios Kopanas, Thomas Leimkühler, and George Drettakis (2023). "3d gaussian splatting for real-time radiance field rendering". In: *ACM Transactions on Graphics* 42.4, pp. 139–1.

Liu, Guilin, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro (2018a). "Image Inpainting for Irregular Holes Using Partial Convolutions". In: *The European Conference on Computer Vision (ECCV), Munich, Germany*.

Liu, Guilin, Kevin J. Shih, Ting-Chun Wang, Fitsum A. Reda, Karan Sapra, Zhiding Yu, Andrew Tao, and Bryan Catanzaro (2018b). "Partial Convolution based Padding". In: *arXiv preprint arXiv:1811.11718*. URL: https://arxiv.org/abs/1811.11718.

Liu, Ligang, Lei Zhang, Yin Xu, Craig Gotsman, and Steven J. Gortler (2008). "A Local/Global Approach to Mesh Parameterization". In: *Proceedings of the Symposium on Geometry Processing, Copenhagen, Denmark*. SGP '08. Copenhagen, Denmark: Eurographics Association, pp. 1495–1504.

Maggiordomo, Andrea, Paolo Cignoni, and Marco Tarini (2023). "Texture Inpainting for Photogrammetric Models". In: *Computer Graphics Forum* 42.6, e14735.

Maggiordomo, Andrea, Federico Ponchio, Paolo Cignoni, and Marco Tarini (2020). "Real-World Textured Things: A repository of textured models generated with modern photo-reconstruction tools". In: *Computer Aided Geometric Design* 83, p. 101943.

Mildenhall, Ben, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng (2021). "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1, pp. 99–106.

Myles, Ashish and Denis Zorin (2012). "Global parametrization by incremental flattening". In: *ACM Trans. Graph.* 31.4, pp. 1–11.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, pp. 234–241.

Sarlin, Paul-Edouard, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk (2019). "From coarse to fine: Robust hierarchical localization at large scale". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12716–12725.

Schonberger, Johannes L and Jan-Michael Frahm (2016). "Structure-from-motion revisited". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113.

Tancik, Matthew, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. (2023). "Nerfstudio: A modular framework for neural radiance field development". In: *ACM SIGGRAPH 2023 Conference proceedings*, pp. 1–12.

Wang, Chen, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu (2022). "Nerf-sr: High quality neural radiance fields using supersampling". In: *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 6445–6454.

2. End2End Avatar Production Pipeline

Alessandro Inguglia[1] *and Ioannis Paraskevopoulos*[1]

[1] ThinGenious PC (THING), Greece

Abstract. This chapter details the development of a fully automatic pipeline for creating hyper-realistic, animatable 3D avatars. The work chronicles a strategic evolution from an initial pipeline reliant on controlled photogrammetric scans to a more accessible and scalable system that generates high-fidelity avatars directly from smartphone videos. We present the final three-stage architecture: (1) photorealistic 3D reconstruction using Neural Radiance Fields (NeRFs) combined with robust Structure from Motion (SfM); (2) non-rigid registration of a proprietary, animation-ready template mesh to the reconstruction, ensuring consistent topology; and (3) automatic generation of the final asset, complete with 4K textures, a full skeletal rig, and 51 ARKit-compatible facial blendshapes. The pipeline achieves a high degree of quantitative resemblance to the human subject and a total processing time of approximately 15 minutes on consumer hardware. This work significantly lowers the barrier to mass avatar production, enabling scalable deployment for different XR applications.

2.1 Introduction

The creation of realistic and interoperable digital human avatars is an important element for immersive Extended Reality (XR) experiences. Recent years have seen remarkable progress in generating 3D human representations from various inputs, including sparse 2D images [Peng et al. 2021; Burkov et al. 2023], single RGB images [Saito et al. 2019; Saito et al. 2020], and monocular videos [Guo et al. 2023; Jiang et al. 2022; Weng et al. 2022]. While these methods have pushed the boundaries of what is possible, challenges related to geometric detail, texture fidelity, and scalability often remain.

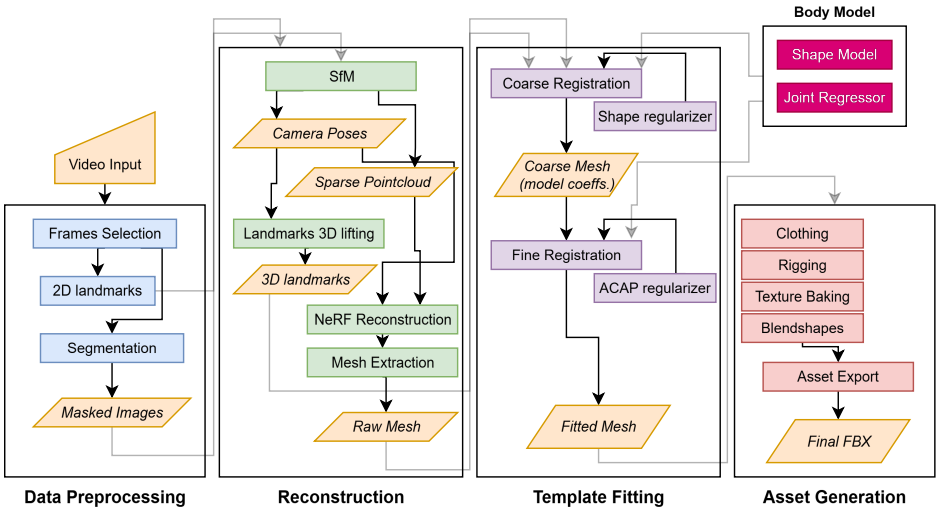


Figure 2.1: Architecture of the avatar pipeline component, showing the data flow from video input to final avatar asset.

Our work began by exploring a pipeline that processed high-quality photogrammetric scans captured in a controlled studio environment. While this method yielded precise geometric accuracy, it presented significant scalability challenges, requiring specialized hardware and user visits to a dedicated facility. To overcome these limitations and democratize the creation of personalized digital twins, a strategic pivot was made towards a more flexible pipeline that leverages consumer-grade technology. This chapter presents the final, fully automatic system that transforms a simple video, captured with a standard smartphone, into a game-ready, rigged, and textured 3D avatar for a broad range of XR applications.

2.2 Methodology and Results

2.2.1 System Architecture

The final video-to-avatar workflow is designed as a modular, multi-stage pipeline that ensures both high-quality results and processing efficiency. The architecture, illustrated in Figure 2.1, is composed of three main sequential stages: (1) Neural Radiance Fields (NeRF)-based reconstruction [Mildenhall et al. 2020], where a photorealistic 3D model of the subject is created from video; (2) template mesh registration, where a standard-

ized template mesh is non-rigidly deformed to match the reconstruction, ensuring a consistent topology for animation; and (3) asset generation, where the final, application-ready avatar is produced with a full set of textures, a skeletal rig, and facial blendshapes. This staged approach allows for a clear separation of concerns, from initial data capture and 3D reconstruction to the final preparation of a functional and interoperable digital asset.

2.2.2 Stage 1: Photorealistic 3D Reconstruction from Video

The first stage of the pipeline is dedicated to reconstructing a high-fidelity 3D model of the subject from a standard video. This process begins with a standardized video capture protocol designed to be easily followed by users with consumer-grade cameras, such as smartphones. The protocol involves recording a 45-60 second video of the subject in an A-pose, ensuring good lighting and a clear view of the entire body, followed by close-ups of the face. From this video, an intelligent frame selection algorithm automatically extracts an optimal set of 120 frames for processing. This selection is not random; it prioritizes frames based on a combination of factors, including the visibility and size of the subject's face, image sharpness (measured via Laplacian variance), and spatio-temporal coherence, ensuring that the subsequent reconstruction is based on the highest quality source material.

Once the frames are selected and the subject is isolated from the background using a neural segmentation network, the pipeline performs Structure from Motion (SfM) to determine the precise 3D position and orientation (pose) of the camera for each frame. For this critical task, we employ the Hierarchical Localization (HLOC) framework [Sarlin et al. 2018], which leverages a combination of deep learning models for robust feature extraction and matching. Global features are extracted using NetVLAD [Arandjelovic et al. 2016] to identify likely image pairs, while local features are detected and described by SuperPoint [DeTone et al. 2018]. These features are then matched across images using SuperGlue [Sarlin et al. 2020], a graph neural network-based matcher that excels at finding reliable correspondences even in challenging conditions. The resulting matches are geometrically verified before being passed to the COLMAP[1] SfM algorithm, which triangulates the camera poses and generates a sparse 3D point cloud of the subject.

The camera poses and processed images serve as input to a NeRF model [Mildenhall et al. 2020], which learns a continuous volumetric representation of the subject. Our pipeline uses an optimized implementation based on Nerfacto [Tancik et al. 2023], which incorporates several advanced techniques for rapid and high-quality reconstruction. These include multi-resolution hash encoding for efficient spatial representation

[1] https://demuc.de/colmap/

Figure 2.2: Render of an evaluation frame from a NeRF dataset. From left to right: occupancy, ground truth, RGB render, depth, normals

Figure 2.3: Comparison of Neural Radiance Field (NeRF) renderings and ground truth (GT) images across multiple viewpoints for two subjects. Each pair shows the NeRF-synthesized view (Left) alongside its corresponding ground truth reference (Right)

[Müller et al. 2022], a two-stage proposal network for adaptive sampling in detailed regions, and a scene contraction method that enables consistent handling of diverse capture environments [Barron et al. 2021]. The resulting rendered scene can be seen in Figure 2.2 and Figure 2.3. Crucially, the reconstruction process is centered and scaled using 3D human pose landmarks detected in the images, which focuses the network's capacity on the subject and accelerates convergence. After a training process of 10,000 iterations, the learned neural representation is converted into a dense, textured 3D mesh using Poisson surface reconstruction, providing the raw geometric data for the next stage of the pipeline.

2.2.3 Stage 2: Template Mesh Registration

The raw mesh extracted from the NeRF model, while geometrically accurate, lacks the consistent structure required for animation and interoperability. The second stage of the pipeline addresses this by performing non-rigid registration of a template mesh to the reconstruction. This template is the basis of the pipeline's animation capabilities, featuring a clean, consistent topology optimized for deformation, a standardized UV map for texturing, a pre-defined skeleton with skinning weights, and a full set of 51 ARKit-compatible facial blendshapes. By fitting this template to each subject, we ensure that every generated avatar shares a common underlying structure, making all subsequent asset generation steps streamlined and reliable.

The fitting process itself is a two-phase optimization designed to robustly align the template to the target scan. The first phase, coarse registration, establishes the overall body shape and pose. This is achieved using a proprietary parametric shape model based on Principal Component Analysis (PCA). Similar to other statistical body models [Loper et al. 2015], our model is trained on a diverse dataset of 3D human scans and can represent a wide variety of body shapes with a compact set of parameters. By optimizing these parameters to minimize the Chamfer distance [Fan et al. 2017] to the target mesh, guided by 3D landmarks detected on the scan, the template is quickly molded into the general shape of the subject.

Following this, a fine registration stage captures subject-specific details that are not represented by the parametric model. This non-rigid deformation directly optimizes the template's vertex positions to precisely match the target surface. To prevent unrealistic stretching and preserve the template's high-quality surface properties, this optimization is constrained by an As-Conformal-As-Possible (ACAP) regularizer [Yoshiyasu et al. 2014]. This regularizer ensures that local transformations are angle-preserving, resulting in a smooth and natural-looking deformation. The final result, as shown in Figure 2.4, is a fitted mesh that perfectly marries the subject's unique geometry with the template's animation-ready topology.

2.2.4 Stage 3: Asset Generation and Finalization

With the fitted mesh in place, the final stage of the pipeline generates the complete, application-ready avatar asset. This involves transferring all the necessary components for appearance and animation from the template to the newly fitted geometry. The first step is to transfer the photorealistic appearance from the NeRF model to the fitted mesh via texture baking. For each vertex on the fitted mesh, a ray is cast towards the NeRF model along the surface normal, and the learned color is sampled. This process effectively "paints" the NeRF's appearance onto the template's standardized UV layout,

Figure 2.4: Results from various steps of the pipeline. First row: SfM pointcloud and cameras and 3D landmarks; Second row: reconstructed mesh and the fitted template mesh

creating a high-resolution 2K or 4K texture map that captures the subject's detailed appearance, from skin tone to clothing texture.

Next, the avatar is made ready for animation. The template's skeleton is adapted to the fitted mesh by first predicting the correct joint positions using a specialized regressor and then transferring the skinning weights, which define how the mesh deforms during movement. Simultaneously, the 51 ARKit-compatible facial expressions are transferred from the template. This is achieved through a deformation transfer [Sumner and Popović 2004] technique, which maps the expressive deformations of the template's face onto the subject's unique facial geometry, enabling nuanced and realistic animation. Finally, all components—the fitted mesh, UV maps, textures, skeleton, and blendshapes—are packaged into a single, interoperable file, either FBX or GLB. This ensures the avatar is immediately usable in a wide array of 3D engines and content creation tools, such as Unity, Unreal Engine, and Blender.

2.2.5 Performance and Evaluation

The performance of the pipeline was evaluated in terms of both processing speed and the quality of the final output. The entire end-to-end process, from video upload to final asset export, completes in approximately 15 minutes on high-end consumer hardware (e.g., NVIDIA RTX 3090/4090). The core template fitting and asset gener-

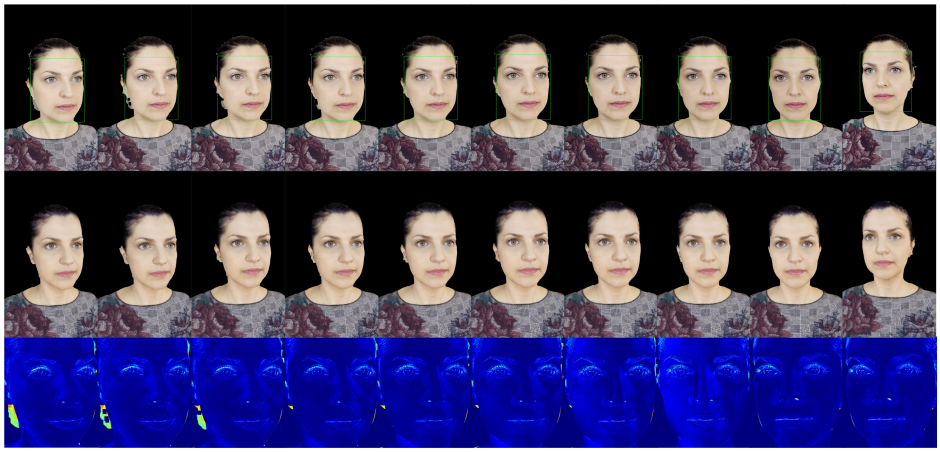


Figure 2.5: During the evaluation step, GT images are compared with the rendered mesh, with focus on the face region.

ation steps are particularly efficient, taking less than 5 minutes to run once the initial NeRF reconstruction is complete. This identifies the neural reconstruction as the main computational bottleneck and a key area for future optimization.

To validate the quality and fidelity of the generated avatars, a comprehensive evaluation framework was developed. This framework renders the final 3D avatar from the same camera viewpoints as the original input video and compares them using a combination of metrics. Pixel-level geometric accuracy is measured using L2 distance on facial regions, while high-level identity preservation is assessed using the DeepFace facial recognition model [Serengil and Ozpinar 2024] to verify that the rendered face is recognized as the same person from the source images (Figure 2.5). In a test on a sample set of avatars, the system achieved a high degree of similarity, with an average L2 distance of 0.077 (a similarity score of 92.3%) and a 100% verification rate from DeepFace. These results provide strong quantitative evidence that the pipeline produces avatars that are not only geometrically accurate but also preserve the unique identity of the subject.

2.3 Conclusions

In this chapter, we have presented the design and evolution of a fully automatic pipeline for the creation of hyper-realistic 3D avatars. The work successfully transitioned from a controlled, photogrammetry-based system to a highly accessible video-based workflow,

democratizing the ability to generate personalized digital twins using only consumer-grade hardware. The final three-stage architecture—combining state-of-the-art neural reconstruction with robust template fitting and asset generation—has proven to be both effective and efficient. The versatility of the generated avatars makes them suitable for a wide range of applications, from clinical rehabilitation and industrial training to social XR platforms.

The significance of this work lies in its contribution to making realistic avatar creation a scalable reality. By removing the dependency on specialized scanning equipment, the pipeline opens the door for mass adoption in a wide range of XR applications. However, the work is not without limitations. The quality of the final avatar remains highly dependent on the quality of the input video, and challenges persist in accurately capturing complex details like flowing hair or intricate clothing. Future research will focus on addressing these challenges, primarily by exploring more advanced neural reconstruction techniques and further optimizing the computational performance of the NeRF stage to bring production times even closer to real-time.

References

Arandjelovic, Relja, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic (2016). "NetVLAD: CNN architecture for weakly supervised place recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5297–5307.

Barron, Jonathan T., Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman (2021). *Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields*. Version Number: 3. URL: https://arxiv.org/abs/2111.12077.

Burkov, Egor, Ruslan Rakhimov, Aleksandr Safin, Evgeny Burnaev, and Victor Lempitsky (2023). "Multi-NeuS: 3D Head Portraits from Single Image with Neural Implicit Functions". In: *IEEE Access*.

DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich (2018). "Superpoint: Self-supervised interest point detection and description". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 224–236.

Fan, Hao, Hao Su, and Leonidas J. Guibas (2017). "A Point Set Generation Network for 3D Object Reconstruction from a Single Image". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Guo, Chen, Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges (2023). "Vid2Avatar: 3D Avatar Reconstruction from Videos in the Wild via Self-supervised Scene Decomposition". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Jiang, Wei, Kwang Moo Yi, Golnoosh Samei, Oncel Tuzel, and Anurag Ranjan (2022). "NeuMan: Neural Human Radiance Field from a Single Video". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 13692 LNCS.

Loper, Matthew, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black (2015). "SMPL: A Skinned Multi-Person Linear Model". In: *ACM SIGGRAPH Asia*.

Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng (2020). "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *European Conference on Computer Vision (ECCV)*.

Müller, Thomas, Alex Evans, Christoph Schied, and Alexander Keller (July 2022). "Instant neural graphics primitives with a multiresolution hash encoding". en. In: *ACM Transactions on Graphics* 41.4, pp. 1–15.

Peng, Sida, Junting Zhang, Qian Li, Wen-na Wang, Jing Liao, Philip H.S. Torr, and Ian Reid (2021). "Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Saito, Shunsuke, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li (2019). "PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization". In: *International Conference on Computer Vision (ICCV)*.

Saito, Shunsuke, Tomas Simon, Jason Saragih, and Hanbyul Joo (2020). "PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Sarlin, Paul-Edouard, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk (2018). *From Coarse to Fine: Robust Hierarchical Localization at Large Scale*. Version Number: 2. URL: https://arxiv.org/abs/1812.03506.

Sarlin, Paul-Edouard, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich (2020). "Superglue: Learning feature matching with graph neural networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4938–4947.

Serengil, Sefik Ilkin and Alper Ozpinar (2024). "Hyper-realistic and context-aware avatars". In: *Journal of Real-Time Image Processing*.

Sumner, Robert W. and Jovan Popović (2004). "Deformation Transfer for Triangle Meshes". In: *ACM SIGGRAPH 2004 Papers*.

Tancik, Matthew, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa (July 2023). "Nerfstudio: A Modular Framework for Neural Radiance Field Development". In: *Special Interest Group on Computer Graphics*

and Interactive Techniques Conference Conference Proceedings. Los Angeles CA USA: ACM, pp. 1–12.

Weng, Chung Yi, Brian Curless, Pratul P. Srinivasan, Jonathan T. Barron, and Ira Kemelmacher-Shlizerman (2022). "HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2022-June.

Yoshiyasu, Yusuke, Wan-Chun Ma, Eiichi Yoshida, and Fumio Kanehiro (Aug. 2014). "As-Conformal-As-Possible Surface Registration". In: *Computer Graphics Forum* 33.5, pp. 257–267.

3. Autonomous 3D Environment Exploration and Reconstruction

Marco Di Benedetto[1] *and Giulio Federico*[1,2]

[1] Institute of Information Science and Technologies, National Research Council (CNR-ISTI), Italy
[2] University of Pisa, Italy

Abstract. This chapter presents an autonomous environment exploration system that integrates Artificial Intelligence (AI)-enhanced 3D reconstruction with intelligent navigation strategies. The system combines SLAM (Simultaneous Localization and Mapping) techniques with Denoising Diffusion Probabilistic Models (DDPM) to generate high-quality 3D reconstructions, compensating for sensor limitations through generative AI. The exploration component employs a multimodal AI agent that integrates computer vision, semantic understanding, and reinforcement learning for autonomous navigation in unknown environments. Experimental validation demonstrates significant improvements in reconstruction quality, with the diffusion model reducing geometric coarseness and correcting acquisition artifacts in 64^3 voxel SDF (Signed Distance Field) volumes. The exploration subsystem combines Proximal Policy Optimization (PPO) reinforcement learning with frontier ranking algorithms, maintaining occupancy grids, semantic grids, and 3D point clouds for context-aware decisions. Performance evaluation shows substantial improvements when integrating semantic understanding with reinforcement learning strategies. The system achieves real-time 3D model correction in under one second and maintains compatibility with standard file formats. Applications include robotics, virtual reality, environmental monitoring, and enhanced spatial accessibility for users with physical limitations.

3.1 Introduction

This chapter presents a comprehensive approach to autonomous environment exploration that combines advanced 3D acquisition techniques with intelligent navigation strategies. The primary objective is to develop a virtual agent capable of autonomously exploring unknown environments while capturing detailed geometric information and generating statistically valid representations of unexplored areas through Artificial Intelligence (AI)-based hallucination techniques.

The solution addresses the challenge of exploring potentially large physical spaces by integrating two complementary technologies: AI-enhanced 3D reconstruction and autonomous exploration algorithms. This dual approach enables comprehensive environment mapping while compensating for sensor limitations and providing enriched spatial understanding for users, including those with physical limitations who may benefit from remote exploration capabilities.

The autonomous exploration system is structured around two interconnected subsystems that work in tandem to achieve comprehensive environment understanding. The first subsystem focuses on 3D acquisition and reconstruction, leveraging advanced generative AI techniques to enhance spatial data quality [Federico et al. 2024]. The second subsystem implements intelligent exploration strategies using reinforcement learning and semantic understanding to guide autonomous navigation decisions.

The integration of these subsystems creates a robust platform capable of operating in diverse environments, from urban driving scenarios to complex architectural spaces. The system's modular design allows for adaptation to various sensor configurations and exploration objectives while maintaining consistent performance across different operational contexts.

3.2 AI-Enhanced 3D Acquisition and Reconstruction

3.2.1 Component Architecture

The 3D acquisition component builds upon Simultaneous Localization and Mapping (SLAM [Cadena et al. 2016]) techniques to generate detailed environmental representations. SLAM technology provides the foundational capability for pose estimation and initial map construction by integrating data from multiple sensors, including cameras, Light Detection and Ranging (LIDAR), and Inertial Measurement Units (IMU). This

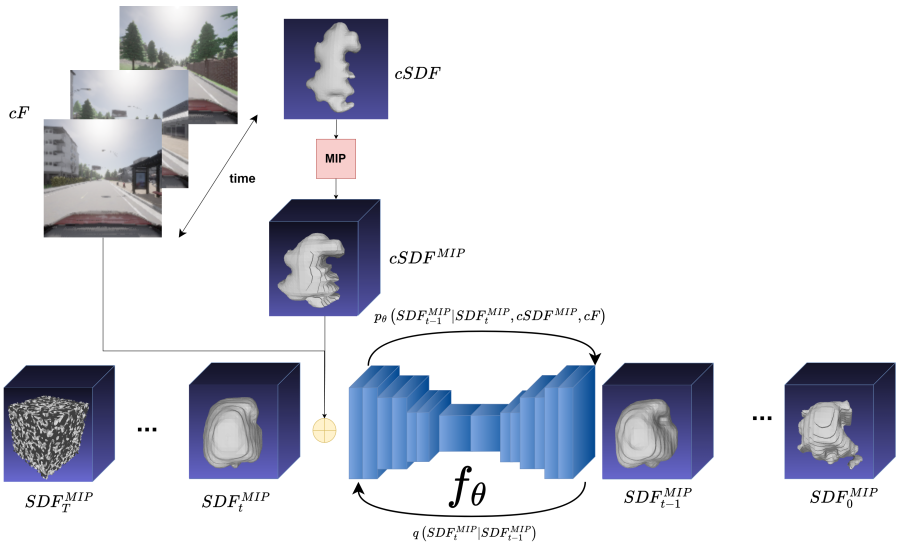


Figure 3.1: Spatio-Temporal Diffusion Model. The diffusion neural network analyzes image sequences and SDF volumes to generate enhanced 3D reconstructions. Figure adapted from [Federico et al. 2024] licensed under CC-BY 4.0.

multi-sensor fusion approach ensures robust spatial understanding even in challenging environmental conditions.

Beyond basic geometric mapping, the system incorporates semantic information processing to create enriched 3D representations. This semantic enhancement enables the agent to understand not only the spatial structure of the environment but also the functional and contextual meaning of different regions and objects within the explored space. An overview of the component architecture is depicted in Figure 3.1.

The core innovation lies in the integration of generative AI, specifically denoising diffusion probabilistic models (DDPM) [Ho et al. 2020], which significantly enhance the quality and completeness of 3D reconstructions. These models address fundamental limitations of traditional 3D scanning approaches, particularly the range constraints of depth cameras and 3D scanners, by generating plausible geometric completions for unobserved or poorly sampled regions.

3.2.2 Generative Reconstruction Process

The reconstruction pipeline processes RGB images alongside image-to-world transformations to generate semantically-enriched 3D models. The diffusion model operates

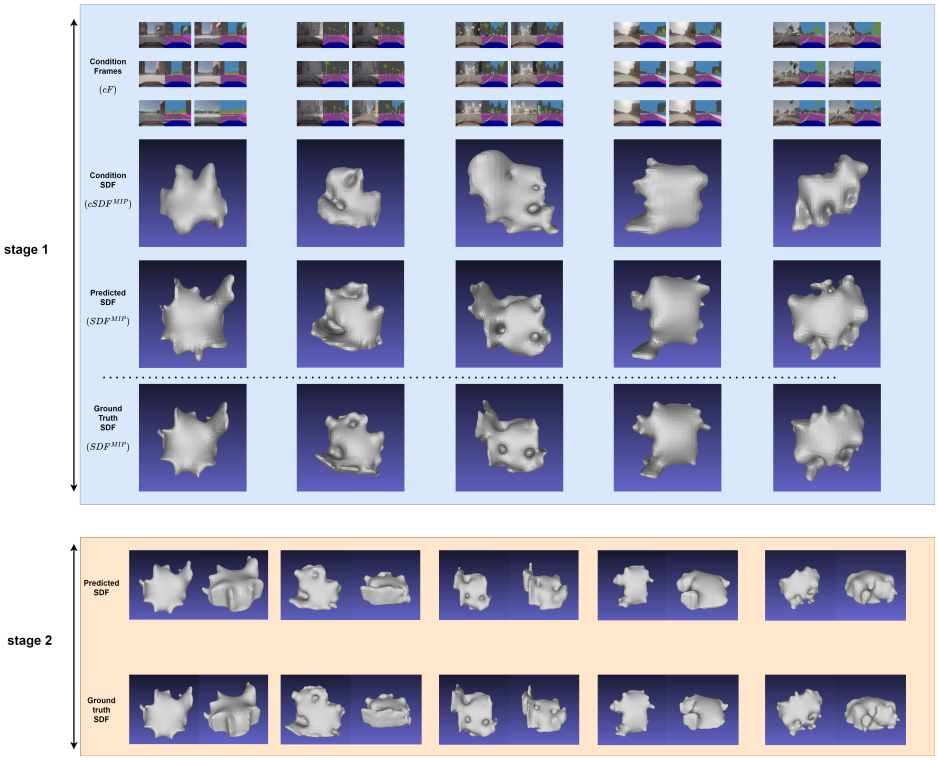


Figure 3.2: Reconstruction results. The two-stage process (generation and rescaling) transforms coarse SDF volumes into detailed scene reconstructions. Figure adapted from [Federico et al. 2024] licensed under CC-BY 4.0.

on signed distance field (SDF) representations [Liu et al. 2024], which provide a mathematically robust framework for 3D geometry representation and manipulation.

The generative process compensates for common acquisition limitations by learning statistical patterns from training data and applying these patterns to fill gaps or correct errors in captured geometry. This approach proves particularly valuable when dealing with incomplete sensor coverage or areas where direct observation is challenging or impossible.

3.2.3 Experimental Results and Validation

System validation was conducted using data from virtual urban driving simulators, providing controlled environments for systematic performance evaluation. Due to compu-

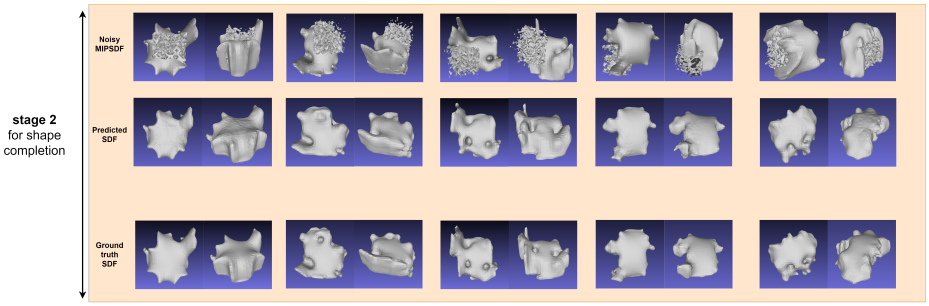


Figure 3.3: Denoising Capability. The diffusion module reconstructs and corrects mesh errors using learned statistical patterns from training data. Figure adapted from [Federico et al. 2024] licensed under CC-BY 4.0.

tational constraints, the generated SDF volumes were limited to 64^3 voxel resolution, which nonetheless demonstrated significant improvements in reconstruction quality.

The experimental results reveal substantial improvements in geometric detail and accuracy compared to raw SLAM output. The diffusion model successfully reduces coarseness in low-resolution input data while maintaining geometric consistency and plausibility (Figure 3.2). Additionally, as shown in Figure 3.3, the system demonstrates the capability in correcting acquisition artifacts and geometric imperfections that commonly arise from sensor noise or systematic measurement errors.

3.3 Autonomous Environment Exploration

3.3.1 Multi-Component AI Agent Architecture

The autonomous exploration capability is implemented through a sophisticated AI agent that integrates multiple machine learning (ML) architectures working in coordination. This multi-modal approach combines computer vision, semantic understanding, and reinforcement learning to create an intelligent exploration system capable of making informed navigation decisions.

As shown in Figure 3.4, the agent architecture consists of several interconnected subsystems, each contributing specialized capabilities to the overall exploration process. At the foundation, the drone platform provides the physical mobility and sensor capabilities necessary for environment interaction. The AI Agent Core Systems contain the primary functional components that enable autonomous decision-making and navigation.

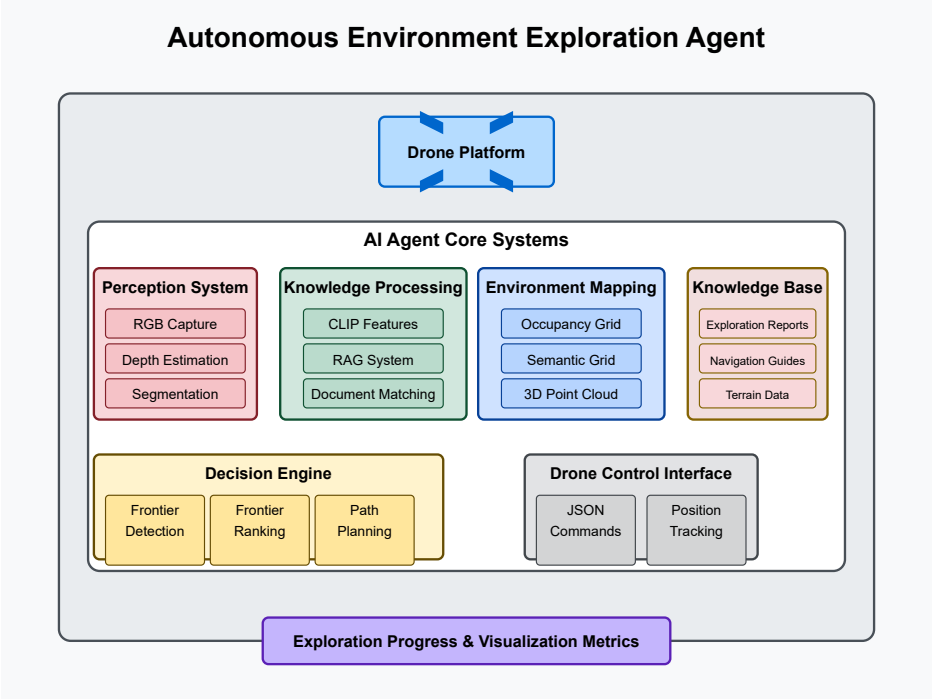


Figure 3.4: Exploration Agent Architecture. The comprehensive system partitions exploration strategy into specialized sub-modules that leverage various ML techniques for coordinated autonomous operation.

The Perception System serves as the sensory interface, processing visual data from onboard cameras to extract RGB imagery, depth estimations, and semantic segmentation of the observed environment. This visual information flows to the Knowledge Processing component, which employs CLIP features for semantic understanding and implements a Retrieval Augmented Generation (RAG) system to match current observations with relevant knowledge from the system's database.

3.3.2 Environment Representation and Mapping

The Environment Mapping module maintains three complementary representations of the explored space. The occupancy grid tracks the spatial distribution of free and occupied areas, providing essential information for navigation planning. The semantic grid associates different regions with object types and functional categories, enabling context-aware exploration decisions. The 3D point cloud reconstruction maintains detailed geometric information about the environment structure.

These representations are continuously updated as new information becomes available through the perception system and knowledge processing pipeline. The multi-layered mapping approach ensures that the agent maintains both detailed local information for immediate navigation decisions and broader contextual understanding for strategic exploration planning.

3.3.3 Decision Making and Navigation Control

The Decision Engine represents the core intelligence of the exploration system, utilizing the comprehensive environmental model to identify frontiers between explored and unexplored regions. The system implements sophisticated frontier ranking algorithms that consider multiple factors, including spatial accessibility, semantic interest, and strategic value for overall exploration objectives.

Path planning algorithms generate optimal routes to selected exploration targets, taking into account both the physical constraints of the drone platform and the informational value of different trajectory options. These high-level navigation decisions are translated through the Drone Control Interface into specific JSON commands that direct the drone's movement system.

3.3.4 Knowledge Integration and Learning

Supporting the core operational components is a comprehensive Knowledge Base containing exploration reports, navigation guides, and terrain data that inform the agent's decision-making processes. This knowledge repository enables the system to leverage prior experience and domain-specific information to improve exploration efficiency and effectiveness.

The system continuously tracks exploration progress and maintains visualization metrics that provide feedback on the effectiveness of different exploration strategies. This monitoring capability enables adaptive behavior and continuous improvement of exploration performance over time.

3.3.5 Reinforcement Learning Integration

The exploration strategy incorporates deep reinforcement learning (RL) through the Unity ML Agents framework, implementing a variant of Proximal Policy Optimization (PPO) for autonomous navigation decisions. The RL component operates within realistic virtual environments that provide accurate physics simulation for drone movement and high-quality visual rendering for computer vision tasks.

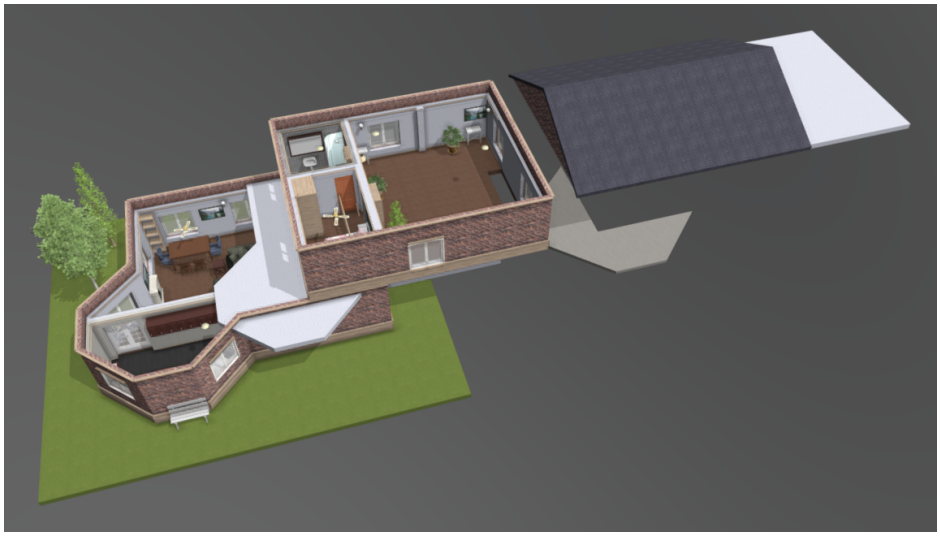


Figure 3.5: Example architectural environment. Complex building structures are used for autonomous discovery algorithm training and evaluation.

Training environments feature architectural buildings with varying complexity levels (Figure 3.5), designed to challenge different aspects of the exploration algorithm. For practical training purposes, doors and windows are removed to eliminate easily avoidable obstacles while maintaining the essential spatial and semantic complexity of real architectural environments.

3.3.6 Performance Evaluation and Results

System performance evaluation employed multiple metrics to assess exploration effectiveness across diverse architectural environments. Each test environment was subdivided into uniform grids to enable precise measurement of exploration coverage and efficiency.

Comparative analysis examined exploration performance with and without multimodal frontier ranking evaluation, demonstrating the value of semantic understanding in guiding exploration decisions. Performance metrics included the percentage of environment explored over time, path efficiency calculated as the ratio of area explored to distance traveled, and time required to achieve 90% environmental coverage.

The experimental results demonstrated substantial performance improvements attributable to the effective integration of semantic understanding with reinforcement learning algorithms. By providing the RL agent with pre-ranked frontier informa-

tion based on multi-modal analysis, the system achieves more efficient exploration while maintaining the flexibility for the agent to develop optimal navigation strategies adapted to specific environmental characteristics.

3.4 System Integration and Applications

The integrated system has been validated through extensive laboratory testing using virtual data from driving simulation environments. The results demonstrate the system's suitability for applications requiring exploration of large physical spaces.

The generative 3D reconstruction component proves particularly valuable in scenarios where comprehensive spatial documentation is required, while the autonomous exploration capability enables efficient coverage of large areas without direct human intervention. This combination makes the system well-suited for applications in environmental monitoring, search and rescue operations, and infrastructure inspection.

Specific performance targets include the capability to propose plausible corrections for visualizing 3D models with missing details in less than one second. The generative reconstruction component and noise-removal features directly contribute to this objective by filling unsampled areas and enhancing 3D asset quality.

3.4.1 Technical Implementation

The current implementation is optimized for 64x64x64 voxel volumes, with configuration parameters allowing adjustment of the number of temporal pairs considered. Dataset statistics are managed through dedicated keys for maximum and minimum values, SDF clipping parameters, mean, and standard deviation values.

Training Pipeline

The training architecture implements a two-stage approach for optimal reconstruction quality. The first stage transforms coarse MIP_SDF volumes into detailed MIP_SDF representations, while the second stage converts MIP_SDF volumes back to standard SDF format. This multi-stage approach proves essential when dealing with datasets containing highly unbalanced distributions of negative and positive values.

Inference System

The complete inference pipeline provides a comprehensive example of system operation from data input through final 3D reconstruction output. This implementation

serves as both a functional tool and a reference for system integration in larger applications.

3.5 Future Directions and Conclusions

The integrated autonomous 3D environment exploration and reconstruction system represents a significant advancement in combining generative AI with intelligent navigation strategies. The system's ability to enhance 3D reconstructions while simultaneously exploring unknown environments opens new possibilities for applications in robotics, virtual reality, and environmental monitoring.

The successful integration of denoising diffusion probabilistic models with SLAM-based reconstruction demonstrates the potential for AI-enhanced spatial understanding. Similarly, the combination of semantic analysis with reinforcement learning for exploration strategies provides a robust foundation for autonomous navigation in complex environments.

Future development directions include expanding the system's capability to handle larger voxel resolutions, implementing real-time processing optimizations, and extending the semantic understanding capabilities to support more diverse environment types. The modular architecture provides a solid foundation for these enhancements while maintaining compatibility with existing implementations and standards.

References

Cadena, Cesar, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose Neira, Ian Reid, and John J. Leonard (Dec. 2016). "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age". In: *IEEE Transactions on Robotics* 32.6, pp. 1309–1332.

Federico, Giulio, Fabio Carrara, Giuseppe Amato, and Marco Di Benedetto (2024). "Spatio-Temporal 3D Reconstruction from Frame Sequences and Feature Points". In: *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops*. IMXw '24. Stockholm, Sweden: Association for Computing Machinery, pp. 52–64.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel (2020). *Denoising Diffusion Probabilistic Models*. arXiv: 2006.11239 [cs.LG].

Liu, Lizhe, Bohua Wang, Hongwei Xie, Daqi Liu, Li Liu, Zhiqiang Tian, Kuiyuan Yang, and Bing Wang (2024). *SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field*. arXiv: 2403.14366 [cs.CV].

4. High-Performance Interactive Streaming

Leesa Joyce[1] *and Mert Ülker*[1]

[1] HOLO-Industrie 4.0 Software GmbH, Germany

Abstract. Immersive technologies such as Augmented and Virtual Reality are central to global digital transformation, yet their adoption is hindered by limited device performance, time-intensive cross-platform development, and growing data privacy concerns. High-performance interactive streaming offers a promising solution by shifting the computational load from mobile Extended Reality (XR) devices to powerful remote servers.

This cross-platform approach enables high levels of detail and responsiveness while reducing development complexity and safeguarding sensitive data. By leveraging remote rendering, developers and organizations can accelerate the creation and deployment of XR applications while providing end users with richer, more secure experiences across devices. In industry use cases such as manufacturing assembly lines, where safety and efficiency are equally important, XR streaming technology plays a very important role. Within the SUN project, XR streaming has been at the core of safety in industry shopfloors and task optimization.

This chapter details the principles of XR streaming, outlines its technical and practical benefits, and presents Hololight's Stream Software Development Kit developed within the SUN project, which demonstrates the capability to stream entire XR applications seamlessly. The results of the project highlight how interactive streaming can overcome current industry challenges, enabling faster development cycles, scalable performance, and safer XR interactions.

4.1 Introduction

The growth of Extended Reality (XR) technologies has expanded opportunities for integrating immersive systems into industrial and entertainment applications. Mobile devices such as smartphones, tablets, and smart glasses have become central to these developments due to their portability and versatility in spatial computing. However, this portability introduces a structural limitation: computational power scales inversely with device mobility. As a result, lightweight XR hardware is unable to deliver the same rendering quality as personal computers (PCs) or dedicated graphics workstations. This imbalance constrains the fidelity of rendered 3D environments and hinders the integration of live-computed data streams, such as point clouds or building information models, into XR applications.

The primary challenge for the XR community is therefore the development of compact devices that can sustain high computational loads while maintaining user mobility. One approach to overcoming this challenge is to outsource rendering tasks to an external infrastructure. Remote rendering and application streaming transfer computationally intensive processes from the XR device to a more powerful server or cloud resource, allowing the device to act primarily as a display and input interface. Such approaches are expected to play a significant role in advancing XR adoption across sectors, including manufacturing, healthcare, education, and training, where real-time visualization of complex data is critical.

Hololight's Hololight Stream (formerly ISAR SDK) represents a state-of-the-art example of a remote rendering solution tailored for XR. By streaming the entire XR application to client devices, Hololight Stream enables developers to bypass the limitations of local hardware. Sensor data from the XR device (e.g., head pose or input gestures) is transmitted back to the server, where rendering takes place. The resulting frames are then streamed to the client using WebRTC protocols. This mechanism supports both cross-platform compatibility and the visualization of large and complex datasets without extensive pre-processing. Importantly, data security is enhanced as sensitive models and application logic remain on secure servers rather than on mobile devices.

XR adoption in industry is constrained by the computational limitations of mobile devices, which form the foundation of many XR systems. Smartphones, tablets, and head-mounted displays lack the GPU and CPU capacity required to sustain complex rendering operations. These limitations manifest in multiple ways:

- *Rendering constraints*: High-polygon models, realistic lighting, and dynamic effects are typically infeasible on mobile XR devices due to hardware performance ceilings;

- *Cross-platform development difficulties*: Heterogeneity across XR devices, engines, and interaction methods necessitates device-specific optimization, increasing development cost and complexity;
- *Data security*: Locally stored data, including confidential industrial models, is vulnerable to loss or theft;
- *Optimization overhead*: Significant effort is required to simplify data models so they can be rendered on mobile platforms. This process can consume the majority of development resources.

The inability of local XR hardware to provide sharp, stable renderings comparable to modern PCs is primarily due to disparities in CPU and GPU performance. Latency in rendering further exacerbates these limitations, producing positional offsets and instability in holographic displays. Although techniques such as reprojection mitigate some effects, latency remains an inherent barrier for XR systems.

To address these issues, several remote rendering solutions have been developed that shift the computational load away from XR devices and onto more powerful cloud or server infrastructure. Microsoft's Azure Remote Rendering allows detailed 3D models to be rendered in the cloud and streamed to headsets such as the HoloLens 2, enabling visualization of complex industrial or medical models without data reduction. Similarly, NVIDIA's CloudXR leverages GPU-based cloud servers to deliver XR applications to lightweight devices. While these solutions demonstrate the feasibility of cloud-based rendering and have been applied across a range of industrial and training scenarios, they typically focus on streaming individual 3D assets or scenes. This approach can lead to separation of application logic from rendered objects, complicating interaction and integration.

Hololight Stream advances beyond these approaches by enabling the streaming of entire XR applications, rather than isolated 3D models. The software integrates directly into Unity-based projects, allowing the server-hosted XR application to transmit pixel streams to client devices while receiving sensor feedback such as head pose. This architecture maintains the cohesion of application logic with visualization, reduces development overhead, and supports high-fidelity rendering of large, complex datasets. By centralizing rendering and data storage on secure servers, Hololight Stream further addresses challenges of data privacy and compliance that arise when sensitive models are deployed directly on mobile XR devices. By streaming entire applications, Hololight Stream enables the visualization and interaction with highly polygonal, data-intensive content such as graphics-intensive 3D objects, 3D Computer-Aided Designs (CAD) models, or Building Information Modelling (BIM) data, which would otherwise be un-

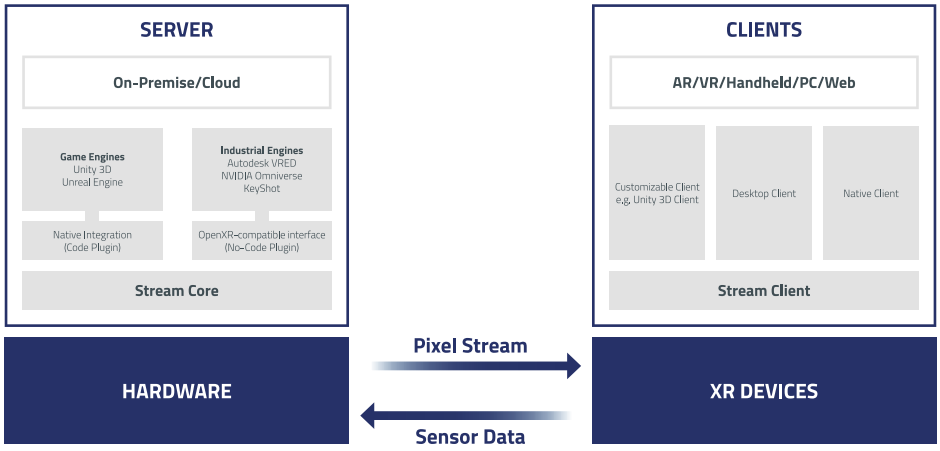


Figure 4.1: Hololight Stream fuctionality

likely on native applications due to the limitation in the processing power of the XR devices.

4.2 Methodology and Results

The Hololight Stream is a Software Development Kit (SDK) that provides a framework for remote rendering in XR systems, enabling real-time transmission of entire AR and VR applications from a remote server to a client device. Instead of executing computationally demanding tasks locally, the rendering process is offloaded to a high-performance server—either on-premises or cloud-based. This approach allows XR devices with limited hardware resources to visualize and interact with complex and data-intensive content.

Once integrated into an XR application, the SDK replaces the need for the application to be deployed directly on the XR device. The application runs on the server, which handles rendering and application logic, while the XR device operates a lightweight client application. This client connects to the remote application and displays the pixel stream it receives (Figure 4.1).

The interaction between client and server follows a bidirectional exchange:

- The client device collects and transmits sensor data, including head and room tracking information, simultaneous localization and mapping (SLAM) inputs, and user interaction data (e.g., gestures);

- The server processes this input in real time, updates the application state accordingly, and performs the rendering task;
- The resulting images are encoded to reduce network bandwidth requirements and transmitted back to the client;
- On the client side, the images are decoded and displayed with minimal delay.

A critical requirement for this architecture is maintaining a low-latency communication loop. In networking terms, this refers to reducing the round-trip time of data packets from client to server and back. To achieve real-time performance, Hololight Stream is designed with latency-optimized protocols and socket-level communication strategies that minimize transmission delays. This ensures synchronization between user actions and visual feedback, which is essential for preserving immersion and usability in XR environments.

4.2.1 Requirements – Hardware, Development and Network

Hololight Stream supports Windows operating systems, specifically Windows 10 or 11. The machine should have at least 16 GB RAM, while 64 GB is recommended for optimal performance. Since Hololight Stream is designed for rendering complex 3D models, a powerful Graphics Processing Unit (GPU) is required for generating the XR frames. It is recommended that the GPU should be at least an NVIDIA GTX 1070Ti for desktop systems. For optimal and improved performance, NVIDIA's RTX series graphics cards are recommended. Please refer to Figure 4.2, Figure 4.3, and Figure 4.4 for the list of requirements.

Security Specifications

Hololight Stream allows users to choose between running it on an on-premises server or a public cloud, giving them control over the storage of sensitive data. During remote rendering, data is never stored on the XR device, enhancing security against device loss or cyber-attacks. A single server transmits data to one client, reducing the risk of interception.

Hololight Stream uses the WebRTC protocol wherein data is transmitted securely with Secure Real-time Transport Protocol (SRTP)[1] and Session Controller Transport Protocol (SCTP)[2] ensuring a robust and safe platform for data transmission.

[1] https://datatracker.ietf.org/doc/html/rfc5764
[2] https://datatracker.ietf.org/doc/html/rfc6083

Component	Minimum Recommendations
Operating System	Windows 10 Windows 11
CPU	Intel Core I7 (9th Generation) AMD Ryzen 7
Cores	6
RAM	16GB
GPU	See Below
VRAM	8GB
Microsoft HoloLens 2	NVIDIA RTX 2080 NVIDIA Quadro RTX 6000

Figure 4.2: Hololight Stream hardware requirements

Component	Specifications
Visual Studio version	2022
Visual Studio components	Game development with Unity; Game development with C++; Desktop development with C++
Unity version	2021.3.x (minimum)
Unity components	Windows Build Support IL2CPP
Mixed Reality Toolkit version	2.8.2, 3.0.0
Graphics card drivers	Latest NVIDIA GTX/RTX graphics card drivers

Figure 4.3: Hololight Stream development environment

Network requirements	
Network	Wi-Fi 6
Network frequency	5 Ghz
Bandwidth	min. 40 Mbit
Round Trip Time	max. 50 ms

Figure 4.4: Hololight Stream network requirements

The Core Technology

Hololight Stream utilizes WebRTC (Web Real-Time Communication) for data transmission. WebRTC is a technology enabling real-time peer-to-peer exchange of audio, video, and data. This communication method relies on encoding, transmitting, and decoding frames in real-time with the aid of software-based video codecs. Due to the latency-sensitive nature of Hololight Stream, it was required to introduce hardware-accelerated video codecs to WebRTC, allowing for faster encode and decode times. Initially, only the H.264 was supported; however, the transition to the H.265 and AV1 codecs was approached to enhance performance and remain state-of-the-art.

H.264, also known as Advanced Video Coding (AVC), is a widely adopted video compression standard that employs block-based motion compensation. It has become the predominant format for recording, compressing, and distributing video, with its usage extending to over 91% of video industry developers by 2019. H.264 remains a versatile and reliable option for video content; however, with larger resolutions and the requirement for increased image quality at lower bit rates, it is no longer suitable as the only codec for Hololight Stream.

H.265, or High-Efficiency Video Coding (HEVC), was developed as a successor to H.264, focusing on enhanced compression capabilities. This codec delivers 25% to 50% better data compression than its predecessor, allowing for the same video quality at lower bit rates or significantly improved quality at equivalent bit rates. It also supports higher resolutions, including 8K UHD, and introduces a Main 10 profile for an increased color depth. While offering substantial bandwidth savings of up to 50%, H.265 requires more processing power, making it more demanding on hardware such as CPUs and GPUs.

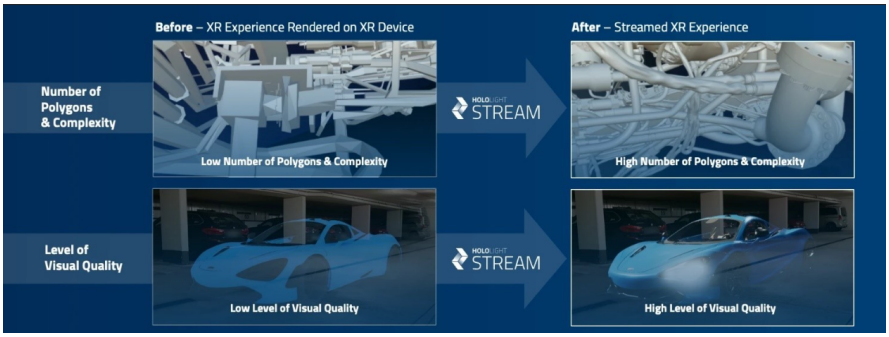


Figure 4.5: Hololight Stream results

4.2.2 Results

After integrating the Hololight Stream Software Development Kit into required applications and having the Hololight Stream client installed on the Microsoft HoloLens2, one is able to successfully stream the entire application on the server. Refer to Figure 4.5. Real-time secure-streaming of the entire Augmented Reality application has been achieved at its original complexity, size, and quality without down-sampling the data. With Hololight Stream, user can run their Unity-built AR application on a powerful workstation, local server, or cloud-based infrastructure.

In addition to enabling high-fidelity streaming, the integration introduces several performance and user experience enhancements. Latency has been significantly reduced through the implementation of predictive algorithms, which optimize data transmission and rendering synchronization. Image quality has been notably improved through advanced encoding and decoding techniques, ensuring clearer and more stable visuals. Furthermore, image stability has been enhanced using head pose prediction, reducing jitter and maintaining alignment between virtual and real-world elements. A gaze tracking mechanism has also been incorporated, allowing for more intuitive user interaction by aligning system responses with the user's visual attention. These improvements collectively contribute to a more immersive, responsive, and seamless AR experience.

The immersive experience requires very low latency in streaming, especially in industry settings. So, significant improvements have been made by developing predictive learning algorithms, anticipating user movements/actions, and allowing the pre-rendering of needed visuals to ensure an optimal user experience.

The image quality has been improved significantly by including a 10-bit color encoding/decoding method.

The devices usually ensure image stability when the application is native (locally hosted on the device). However, with streaming, the device cannot have an impact on image stability and this needed to be addressed. For Hololight Stream, the image stability was improved through head pose prediction. Enabling the feature impacts the server's CPU but ensures that the object remains at a stable location in all directions, even when the user is moving around the object.

Hololight Stream sends the sensor data from the glasses related to eye movement to the server, which can be used as an input method in addition to the pinch gesture.

4.3 Conclusions

The development of high-performance interactive streaming of 3D data through the SUN project marks a significant step toward unlocking the full potential of spatial computing. By leveraging Hololight's expertise in remote rendering, the project demonstrates how XR streaming can overcome the limitations of mobile and lightweight devices, enabling complex AR and VR applications without sacrificing performance or security. Hololight Stream's device-agnostic, server-based approach not only reduces development effort but also ensures scalability, data protection, and ease of integration for future XR solutions. This capability positions Hololight as a key enabler for next-generation immersive technologies, accelerating global digitization and paving the way for more powerful, flexible, and secure XR applications.

5. Open-Vocabulary Understanding of Objects and Scenes

Fabio Carrara[1], *Lorenzo Bianchi*[1], *Nicola Messina*[1], *Claudio Gennaro*[1], *and Fabrizio Falchi*[1]

[1]Institute of Information Science and Technologies, National Research Council (CNR-ISTI), Italy

Abstract. This chapter addresses the challenge of enabling Extended Reality (XR) systems with a flexible, human-like understanding of their surroundings. We present a suite of technologies for open-vocabulary understanding of objects and scenes, moving beyond the limitations of traditional, fixed-category computer vision. The core of our contribution is twofold. The first tackles Open-Vocabulary Detection (OVD) and Segmentation (OVS) to detect and segment objects based on dynamic natural language or visual queries. The second tackles Semantic 3D Scene Reconstruction that builds interactive, queryable 3D models by embedding rich semantic features from vision-language models (CLIP, DINOv2) directly into the scene geometry. The work presented provides a robust framework for developing adaptive, context-aware Augmented Reality (AR) solutions that can significantly enhance human-machine collaboration in complex environments.

5.1 Introduction

The integration of Augmented Reality (AR) into industrial settings holds immense promise for improving efficiency, safety, and training. However, to be truly effective, AR systems must be able to perceive and understand the physical world with a high degree of flexibility. Most conventional systems rely on computer vision models trained to recognize a predefined, fixed set of objects, rendering them brittle and unable to adapt to the dynamic and ever-changing nature of a factory floor or logistics hub.

This chapter introduces a framework for Open-Vocabulary Understanding of Objects and Scenes, designed to overcome these limitations. The goal is to equip AR applications with the ability to recognize and interact with a virtually unlimited range of objects and concepts defined at runtime. We present two core technologies that form the foundation of this framework:

- *Open-Vocabulary Detection and Segmentation (OVD/S)* for flexible 2D object perception (Section 5.2);
- *Semantic 3D Scene Reconstruction* that creates rich, interactive 3D world models imbued with semantic meaning (Section 5.3).

These technologies leverage the power of modern vision-language and self-supervised models (including OWL-ViT, CLIP, and DINOv2) to interpret scenes based on natural language descriptions or visual examples, providing a blueprint for the next generation of context-aware industrial AR systems.

We base our framework on representations learned by visual and multimodal foundation models; the idea of embedding and matching data in a common multimodal space is key for our development in methods that can understand and query objects and scenes in natural language in a training-free manner.

5.2 Open-Vocabulary Object Detection

To enable flexible perception, we investigated Open-Vocabulary Detection (OVD), capable of identifying objects in RGB images based on user-defined vocabularies provided at inference time. This sidesteps the need for retraining for every new object class.

5.2.1 Fine-grained Open-Vocabulary Detection

Supporting fine-grained object descriptions in open-vocabulary detection is a main goal to strive for domain generalization. We devised the FG-OVD (Fine-Grained Open-Vocabulary Detection) benchmark [Bianchi et al. 2024b] to specifically probe the ability of state-of-the-art open-vocabulary object detectors to understand fine-grained properties of objects and their parts. This benchmark addresses a critical limitation of standard open-vocabulary evaluations, which often do not sufficiently test a model's capacity to discern subtle differences between objects, especially in the presence of visually similar "hard-negative" classes. The evaluation protocol of FG-OVD is based on

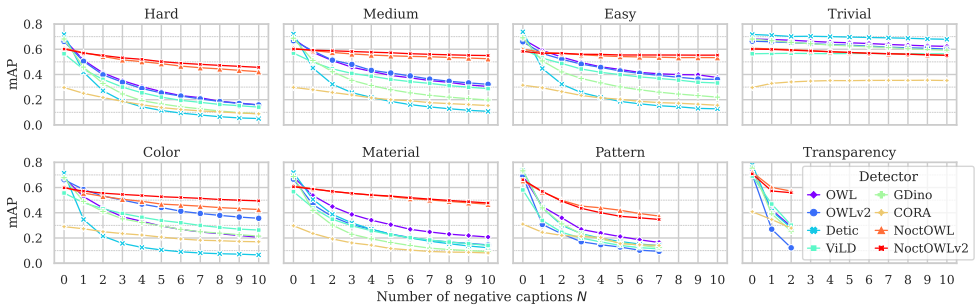


Figure 5.1: Benchmarking state-of-the-art open-vocabulary detectors when increasing the number of negative captions when discerning fine-grained details. We report the mAP varying the number N of negative captions for the different probed detectors and for different categories of attributes to be discerned.

Table 5.1: Performance of image-text matching models in Fine-grained (FG-OVD) and Coarse-grained (COCO) settings. We analyze the performance of image-text matching models. CLIP B/16 represents the standard image-text matching using frozen CLIP embeddings. The other lines show the performance when adding a linear layer on top of the visual representation (rows 2-3), the textual representation (rows 4-5), or both (rows 6-7). The new layers are warmed up on COCO (rows 2, 4, and 6) and subsequently fine-tuned on the fine-grained dataset FG-OVD (rows 3,5, and 7). In the fine-tuned configuration (+FG-OVD rows), we denote the delta between these results and those obtained during the warm-up in parentheses.

	FG-OVD	COCO Retrieval (T→I)		
	Median Rank ↓	R@1 ↑	R@5 ↑	R@10 ↑
CLIP B/16	2.98	22.6	44.1	54.9
Linear (visual only)	3.53	35.4	64.2	75.3
+FG-OVD	1.54(-1.99)	34.3(-1.1)	62.9(-1.3)	74.1(-1.2)
Linear (text only)	3.48	36.0	64.3	75.6
+FG-OVD	1.57(-1.91)	34.7(-1.3)	63.2(-1.1)	74.6(-1.0)
Linear (both)	3.78	**37.2**	**65.6**	**76.6**
+FG-OVD	1.46(-2.32)	35.6(-1.6)	63.9(-1.7)	75.0(-1.6)

dynamic vocabulary generation, where the difficulty increases with the number of negative captions in the vocabulary. The benchmark suite includes different configurations that test various properties like *color, pattern, and material*.

Key findings, published in [Bianchi et al. 2024b], demonstrated that while existing open-vocabulary models perform well in standard benchmarks, they often struggle to accurately capture and distinguish finer object details when faced with a high number of fine-grained negative captions. For instance, models correctly detect objects in the absence of negative captions, but their mAP (mean Average Precision) significantly decreases as the number of fine-grained negative captions increases across various configurations (Figure 5.1). The *Trivial* benchmark, which lacks fine-grained differences between positive and negative captions, showed less degradation, suggesting that the primary challenge lies in correctly classifying object attributes rather than merely localizing objects from complex natural language descriptions. We confirmed this hypothesis in [Bianchi et al. 2024a], where we showed that the performance using groundtruth object locations (i.e., disregarding any errors that detectors may introduce in localization) does not differ meaningfully. Moreover, we observed that fine-grained information needed to discern attributes is present in CLIP, the main backbone of open-vocabulary detectors. We confirmed this by training linear layers over frozen CLIP representations and observing an increased separability of fine-grained attributes while maintaining coarse-grained performance (Table 5.1). Using a weakly labeled training dataset that follows the FG-OVD data preparation pipeline, we contributed new OVD models (NoctOWL and NoctOWLv2 in Figure 5.1) with an improved trade-off between fine-grained and coarse-grained detection capabilities [Bianchi et al. 2025].

5.2.2 OVD Component

In the SUN project, we built a component around the OWL-ViT model, incorporating several key features to enhance its utility in industrial settings:

- *Hierarchical and Visual Vocabularies*, as the system supports complex, nested queries (e.g., find a "person", then a "helmet") and accepts *visual vocabularies*, where classes are defined by representative images;
- *Optimized Real-Time Inference*, as the models are quantized and accelerated with NVIDIA's TensorRT, achieving the low latency required for interactive applications like processing live video from AR headsets;
- *Flexible Integration*, as a RESTful HTTP API ensure the component can be easily deployed and connected to other systems.

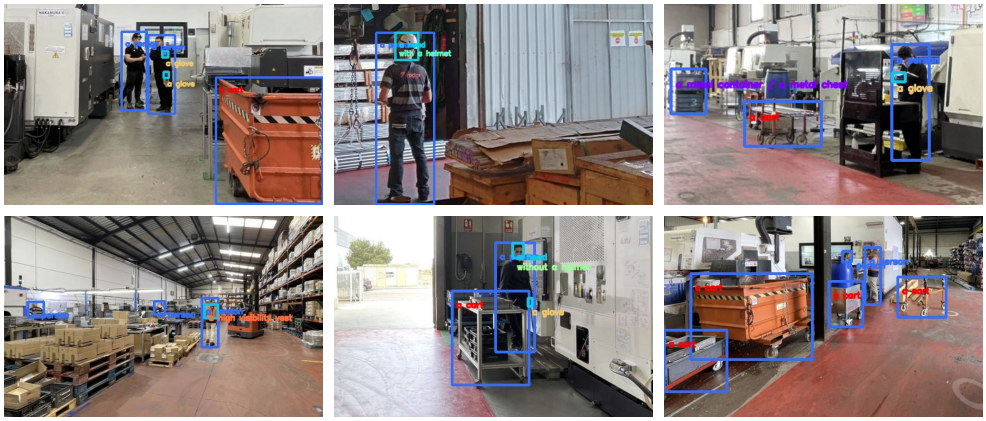


Figure 5.2: Sample predictions of the OVD component on egocentric pictures of the Pilot 2 location (partner FACTOR's shopfloor). Vocabulary: a person, a head (with a helmet, without a helmet), a glove, a high visibility vest, a cart, a metal chest / container.

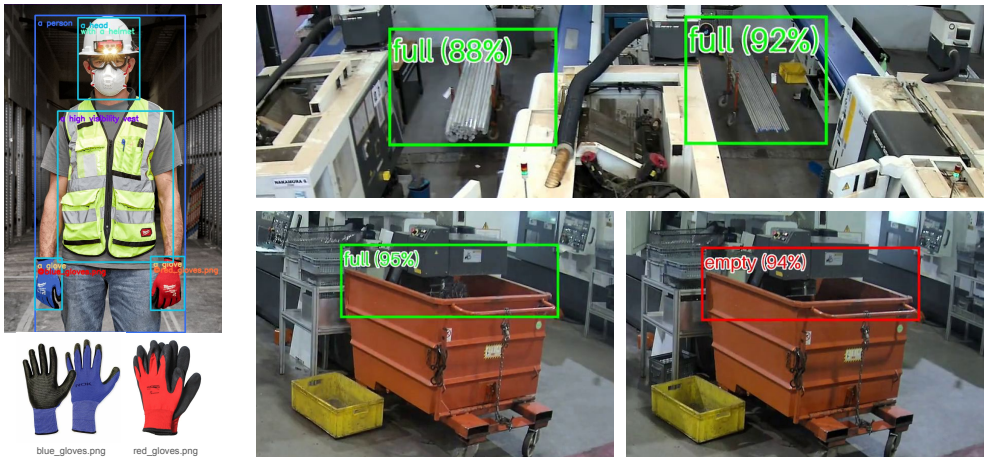


Figure 5.3: Application of the OVD component in the Pilot 2 industrial scenario. (Left) PPE detection using a mixed textual-visual and hierarchical vocabulary. (Right) Container status detection using visual vocabularies to classify material and waste containers.

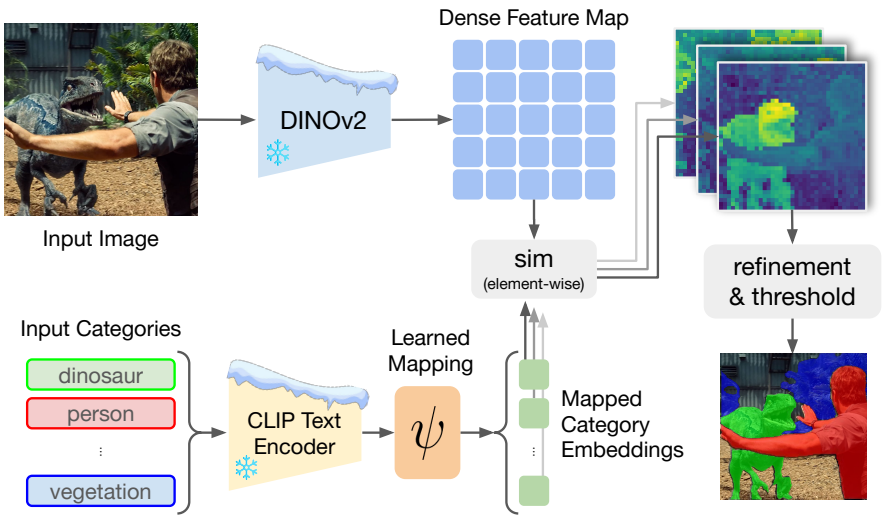


Figure 5.4: Talk2DINO Overview. We align the textual CLIP space to DINOv2's one via a learned mapping function, obtaining fine-grained visual embeddings that can be matched locally against language to provide semantic segmentation. Image from [Barsellotti et al. 2025].

The OVD's capabilities were proven in industrial scenarios, during the SUN Pilot 2 validation (Chapter 21). For workers' safety, OVD processes live video to detect Personal Protective Equipment (PPE) using hierarchical text prompts (Figure 5.2). For logistics, visual vocabularies are used to monitor container status (Figure 5.3), distinguishing between container types and their fill levels.

5.2.3 Open-Vocabulary Segmentation

We also investigated Open-Vocabulary Segmentation (OVS), which, in addition to OVD, also provides pixel-level masks of objects of interest. Specifically, we contributed to *unsupervised* OVS, in which we assume no mask-caption training data is available. We developed a new model for unsupervised OVS named Talk2DINO [Barsellotti et al. 2025] by mapping two foundation models: CLIP [Radford et al. 2021] and DINOv2 [Oquab et al. 2023]. We mapped frozen textual CLIP embeddings to the space of DINOv2 visual patches by learning a simple projection function, i.e., a linear layer or a 2-layer MLP (Figure 5.4). We demonstrated on several segmentation benchmarks (Table 5.2) that simply matching dense DINOv2 features with these mapped CLIP textual features provides good mask candidates, setting a new state of the art in unsupervised OVS.

Table 5.2: Unsupervised Open-Vocabulary Segmentation. Performance (mean IoU) against several state-of-the-art models on multiple segmentation datasets.

Model	Visual Backbone	Frozen	VOC	Context	Stuff	Cityscapes	ADE	Avg
GroupViT [Xu et al. 2022]	Custom ViT	✗	81.5	23.8	15.4	11.6	9.4	28.3
ReCo [Shin et al. 2022]	CLIP	✗	62.4	24.7	16.3	22.8	12.4	27.7
TCL [Cha et al. 2023]	CLIP	✗	83.2	33.9	22.4	24.0	17.1	36.1
MaskCLIP [Zhou et al. 2022]	CLIP	✗	74.9	26.4	16.4	12.6	9.8	28.0
SCLIP [Wang et al. 2024]	CLIP	✗	83.5	36.1	23.9	34.1	17.8	39.1
NACLIP [Hajimiri et al. 2025]	CLIP	✓	83.0	38.4	25.7	38.3	19.1	40.9
FreeDA [Barsellotti et al. 2024]	CLIP+DINOv2	✓	85.6	43.1	27.8	36.7	22.4	43.1
Talk2DINO	DINOv2	✓	**87.0**	**43.5**	**30.3**	**40.8**	**23.5**	**45.0**

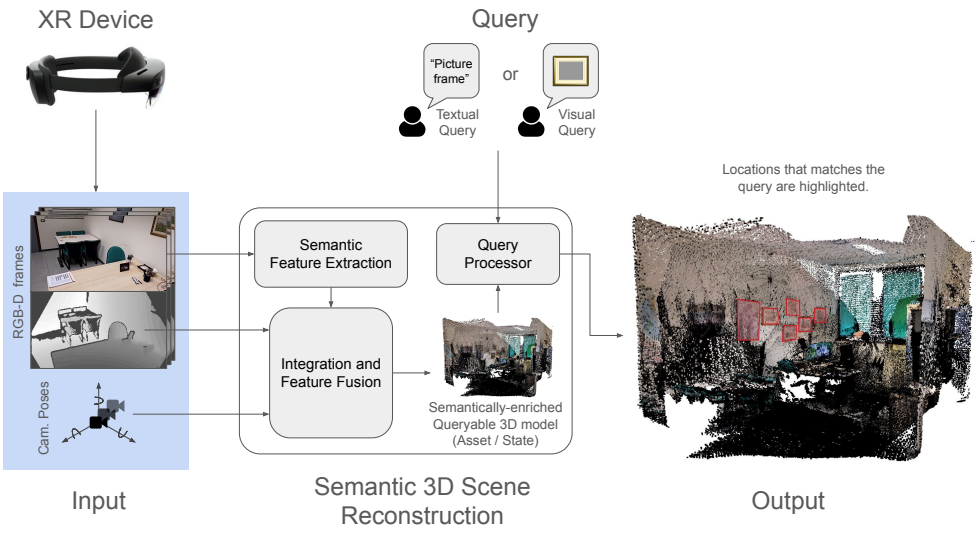
5.3 Semantic 3D Scene Reconstruction

For a more holistic understanding of an environment, we developed a Semantic 3D Scene Reconstruction component. It goes beyond creating simple geometric models by generating detailed, interactive 3D scenes that can be queried semantically.

The process (Figure 5.5, Top) begins by fusing RGB-D video and camera poses into a point cloud. The next step comprises semantic enrichment: RGB frames are segmented using the Segment Anything Model (SAM), and feature vectors from CLIP and DINOv2 are extracted for each segment. These features are then projected onto the 3D points, creating a point cloud where each point carries rich semantic data (see Figure 5.6).

We further refined the query process with a multi-feature fusion mechanism. An initial search with global CLIP features provides a coarse localization, which is then sharpened using more discriminative DINOv2 features. This significantly reduces noise and improves query accuracy (Figure 5.5, Bottom), allowing a user to accurately find an object like a "picture frame" within the reconstructed scene using a simple text prompt.

As a parallel activity, we also explored a full 3D pipeline (not relying on 2D foundation models, but only on 3D ones) to identify objects and extract their semantics. The findings, published in [D'Orsi et al. 2025], show major limitations of the available 3D foundation models and trace a path for future research in this direction.



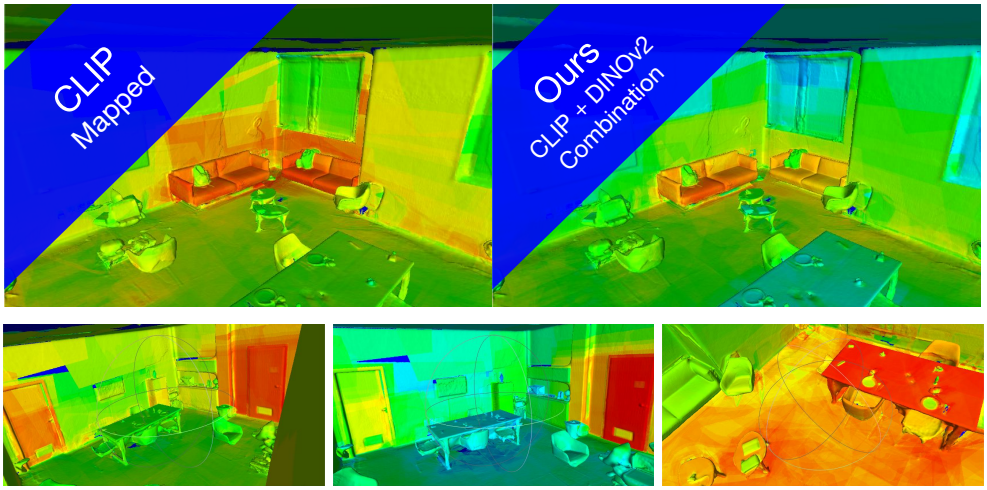


Figure 5.5: On the top, an overview of the semantic reconstruction pipeline. On the second and third row, a qualitative comparison of query results ("couch" and "door"), showing improved localization and reduced noise when using multi-feature fusion (Right) versus a single-feature query (Left). The rightmost image in the last row show limitations (segmentation inaccuracies) still to be tackled.

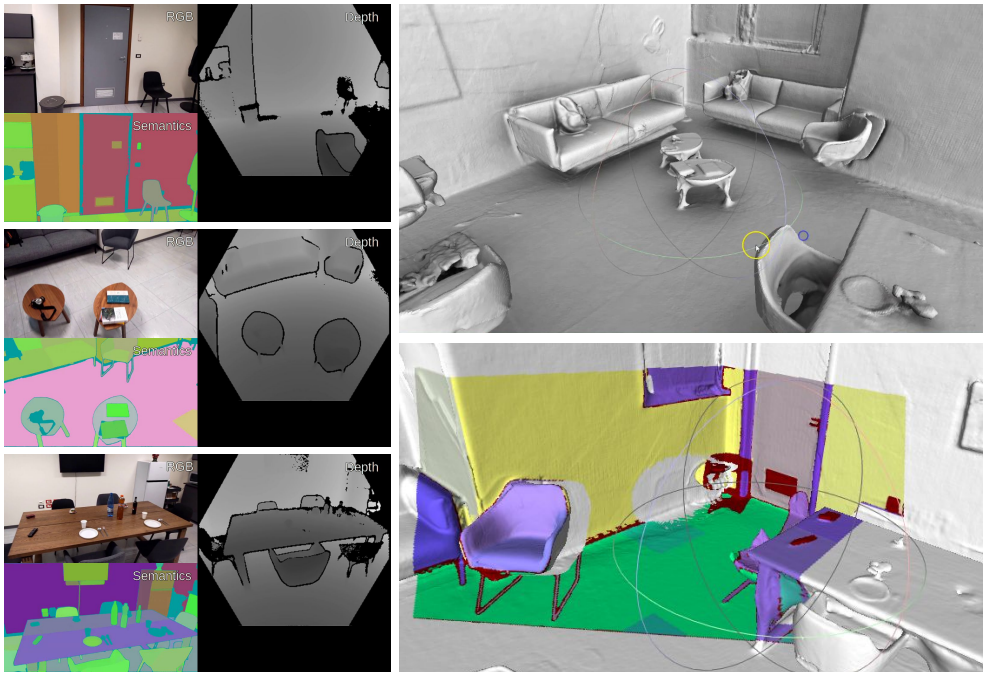


Figure 5.6: Intermediate outputs of the Semantic 3D Reconstruction pipeline. (Left) Three input data points (RGB and Depth) and the semantic encoding of recognized objects (as 3-component PCA of the CLIP features associated with each pixel) for an indoor walk done with a HoloLens 2 device. (Right) Two viewpoints of the coarse reconstructed 3D scene.

5.4 Conclusions

We have presented a framework for open-vocabulary understanding of objects and scenes based on Open-Vocabulary Detection and Semantic 3D Scene Reconstruction, providing the foundational perceptual capabilities for creating intelligent AR systems.

While this framework marks a significant step forward, certain limitations exist. Open-vocabulary Detection and Segmentation on images can still improve their domain generalization to compete with fine-tuned solutions in specific domains. The Semantic 3D Reconstruction component remains experimental and requires further improvements for deployment in uncontrolled environments. Additionally, the system's performance is ultimately dependent on the underlying capabilities and potential biases of the foundation models it employs. Several novel and competitive technologies for 3D processing and understanding need to be evaluated and compared to foster the development of new, efficient, and effective representations for static and dynamic environments.

References

Barsellotti, Luca, Roberto Amoroso, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara (2024). "Training-Free Open-Vocabulary Segmentation with Offline Diffusion-Augmented Prototype Generation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Barsellotti, Luca, Lorenzo Bianchi, Nicola Messina, Fabio Carrara, Marcella Cornia, Lorenzo Baraldi, Fabrizio Falchi, and Rita Cucchiara (2025). "Talking to DINO: Bridging Self-Supervised Vision Backbones with Language for Open-Vocabulary Segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Bianchi, Lorenzo, Fabio Carrara, Nicola Messina, and Fabrizio Falchi (2024a). "Is CLIP the main roadblock for fine-grained open-world perception?" In: *2024 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, pp. 1–8.

Bianchi, Lorenzo, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi (2024b). "The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22520–22529.

Bianchi, Lorenzo, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi (2025). "Fine-grained Open-vocabulary Object Detection". In: *Under Review*. URL: https://doi.org/10.5281/zenodo.17339026.

Cha, Junbum, Jonghwan Mun, and Byungseok Roh (2023). "Learning To Generate Text-Grounded Mask for Open-World Semantic Segmentation From Only Image-Text Pairs". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

D'Orsi, Domenico, Fabio Carrara, Fabrizio Falchi, and Nicola Tonellotto (2025). "Breaking the 2D Dependency: What Limits 3D-Only Open-Vocabulary Scene Understanding". In: *2025 International Conference on Content-Based Multimedia Indexing (CBMI)*. In Press.

Hajimiri, Sina, Ismail Ben Ayed, and Jose Dolz (2025). "Pay attention to your neighbours: Training-free open-vocabulary semantic segmentation". In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 5061–5071.

Oquab, Maxime, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. (2023). "DINOv2: Learning Robust Visual Features without Supervision". In: *arXiv preprint arXiv:2304.07193*.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. (2021).

"Learning Transferable Visual Models From Natural Language Supervision". In: *International conference on machine learning*.

Shin, Gyungin, Weidi Xie, and Samuel Albanie (2022). "ReCo: Retrieve and Co-segment for Zero-shot Transfer". In: *Advances in Neural Information Processing Systems*.

Wang, Feng, Jieru Mei, and Alan Yuille (2024). "SCLIP: Rethinking Self-Attention for Dense Vision-Language Inference". In: *European conference on computer vision (ECCV)*.

Xu, Jiarui, Shalini De Mello, Sifei Liu, Wonmin Byeon, Thomas Breuel, Jan Kautz, and Xiaolong Wang (2022). "GroupViT: Semantic Segmentation Emerges From Text Supervision". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhou, Chong, Chen Change Loy, and Bo Dai (2022). "Extract Free Dense Labels from CLIP". In: *European conference on computer vision (ECCV)*. Springer, pp. 696–712.

Collaborating and Interacting

This part introduces technologies that enhance communication, perception, and engagement within extended reality environments. Wearable haptic systems providing thermal, tactile, and kinesthetic feedback enrich sensory immersion and user awareness. Gaze-based interaction enables seamless, hands-free collaboration, while task optimization tools integrate XR visualization and intelligent prioritization to support efficient teamwork and decision-making. Together, these technologies create a more effective and natural perception of the XR environment and a more responsive and collaborative human interaction.

6. Thermal Feedback in Wearable Haptics

Jonathan Muheim[1], *Mélina Lasfargues*[1], *Aline Brunner*[1], *Solaiman Shokur*[1], *and Silvestro Micera*[1]

[1] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Abstract. Thermal sensations play a vital role in enriching our tactile experiences, yet temperature feedback remains an underexplored modality in virtual reality (VR) systems. This chapter presents the development and evaluation of a wearable thermal feedback device designed to deliver precise warm and cold stimuli as part of the SUN project. After reviewing state-of-the-art technologies for cutaneous thermal stimulation, we describe the design rationale, technical implementation, and evolution of our system—from early prototypes to a fully wearable solution capable of operating up to two independently controlled thermodes. Leveraging thermoelectric modules and passive heatsinks, the device ensures safe and accurate temperature delivery across a 15–42 °C range. We then detail an experimental study using the device to investigate thermal illusions, demonstrating that modulating temperature intensity between two stimulation points can shift the perceived location of thermal sensations. Results show that warm stimuli tend to integrate more readily into single-point sensations, while cold stimuli more effectively support spatial differentiation. These findings highlight both the technical viability and perceptual richness of temperature-based feedback, offering promising directions for multisensory VR experiences.

6.1 Introduction

From grasping a cold soda can to holding someone's hand, thermal sensations are an integral part of our daily lives. Every object we touch—whether cold, neutral, or warm—contributes to our overall sensory experience through its thermal quality. These cues enrich our tactile perception and give rise to complex somatosensory experiences.

Within the SUN project, we aim to develop an innovative wearable device capable of delivering realistic temperature feedback. This chapter outlines current state-of-the-art technologies in cutaneous thermal stimulation, introduces our developed system, and demonstrates how it can be used to investigate nuanced thermal perceptions in immersive environments.

While significant progress has been made in replicating touch, vibration, and force in VR through various haptic interfaces, temperature feedback has received comparatively less attention. Simulating heat is relatively straightforward using resistive elements; however, reproducing cold sensations—especially lowering skin temperature below its baseline—presents notable challenges.

From a neurophysiological perspective, thermal perception relies on distinct thermoreceptors that respond preferentially to either cooling or warming stimuli, typically relative to a baseline skin temperature of 30–34°C [Filingeri 2016]. This biological asymmetry is mirrored in actuator technology. While generating warmth can be achieved with basic components such as electrical resistors, eliciting both cold and warm sensations requires more complex systems. Existing thermal devices generally follow one of two approaches: (1) delivering controlled-contact fluids (e.g., mixing hot and cold water or air), or (2) generating a direct heat flow through the skin surface, typically by extracting heat to induce cooling.

The first type of device typically relies on the presence of two fluids (e.g., water or air) at defined temperatures and a system to mix them in a controlled manner to reach the desired stimulation temperature (e.g., [Cai et al. 2020]). The need for pumps or valves and tanks to store the liquids is a real challenge in making lightweight portable devices. On the other side, the second type of device relies on thermoelectric modules, which act as heat pumps. When an electrical current passes through them, a heat flow is generated between their two sides, resulting in the extraction of heat from one side and its transfer to the other. More practically, this leads to one side cooling down while the other heats up. Interestingly, when inverting the current, the direction of heat transfer is also inverted. These devices can be very thin and lightweight, and are available in a variety of dimensions, which explains their extended use for wearable feedback devices. For the system to provide continuous cold sensations, the heat extracted from the skin

must be efficiently dissipated from the hot side of the Peltier element. An effective approach is based on the attachment of a heatsink to the thermoelectric module. It might be noted that across the different topologies of heatsinks, passive heatsinks seem to be the most appropriate one in this specific use case. Indeed, while active heatsinks usually allow better heat dissipation, the additional power consumption and the possible tactile cues generated by the use of a fan make their integration extremely challenging.

The investigation of literature about thermal feedback devices highlights a variety of devices to provide thermal stimulation of the skin that vary mainly in the size and location of stimulation and the type of technology used. However, the strict constraints on weight and size accompanying wearable feedback devices make an actuator based on the thermoelectric effect (i.e., Peltier elements) the most suited candidate.

The thermal feedback devices found in the literature are designed to provide temperature stimuli on specific locations of the body: on the forearm [Günther et al. 2020], the fingers [Lee et al. 2020], the foot sole [Gallo et al. 2014], or even the face [Peiris et al. 2017]. This approach provides limited flexibility in the location where the stimuli are applied to the body. We anticipate that, depending on the type of interaction occurring in the virtual environment, one body area would be more suited to target of thermal feedback than the others, and that this choice of location could change depending on the scenario. Hence, designing a device that allows placement on different body parts would be beneficial to render a wider range of thermal interaction. Additionally, individual differences in thermal sensitivity are likely to arise in the last pilot involving patients with sensorimotor impairments.

The development of the initial series of prototypes provided valuable insights into refining design specifications and identifying key challenges in building the thermal feedback component. Using off-the-shelf electronic modules and a commercially available temperature controller, the first prototype—shown in Figure 6.1—was constructed. This version integrated the control electronics, power supply, and an interface for a custom wearable thermal display. It successfully delivered temperature stimuli within the desired range (15–42°C) and was capable of replicating the thermal signatures of materials such as copper, glass, and plastic [Iberite et al. 2023; Muheim et al. 2024]. It was additionally used to create an illusory contact with wet objects [Ploumitsakou et al. 2024]. These tests validated the overall system architecture. However, prolonged use of the prototype revealed limitations, particularly in response time. A delay of several seconds between stimulation onset and temperature stabilization emphasized the need for custom-designed control electronics. Such enhancements would not only improve performance but also enable a more compact and integrated design by consolidating all components onto a single Printed-Circuit-Board PCB. This initial prototype

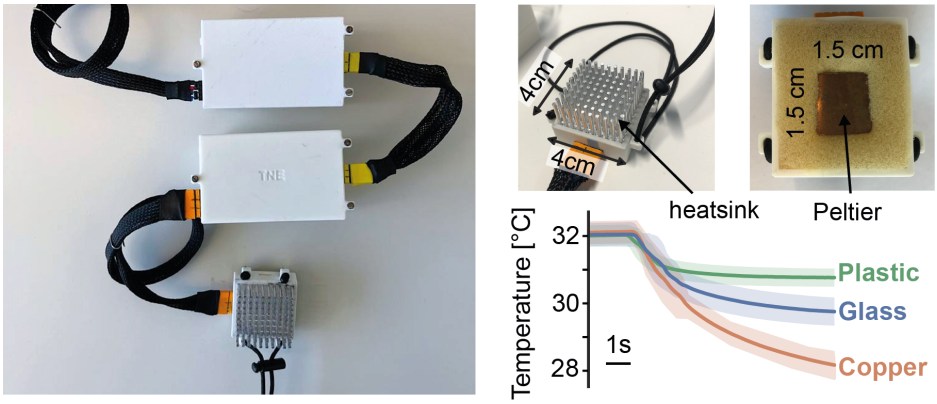


Figure 6.1: The first fully thermal feedback prototype (Left) is composed of two boxes housing the control electronics and the battery, and a thermal display (Top Right) providing the thermal stimuli onto the skin of the wearer. It was successfully used to reproduce the thermal signature of three materials (Bottom Right).

also demonstrated the feasibility of a fully portable system, setting the stage for future miniaturization efforts.

Building on these foundations, the second iteration of the thermal feedback device has undergone substantial advancements. It is fully wearable, featuring a compact control unit capable of simultaneously operating up to two thermal displays (Figure 6.2). Each display is independently regulated by a Proportional-Integrative-Derivative (PID) controller, ensuring precise and safe thermal stimulation within the 15–42°C range. The stimulation sites, each measuring 15 × 15 mm, are comfortably secured to the forearm using elastic bands, allowing for stability during experimental use. Additionally, an integrated battery powers the system for up to 10 hours, supporting extended, untethered operation in realistic usage scenarios.

6.2 Investigation of Thermal Illusions

The control of two thermodes allows investigating the creation of thermal illusions. The first thermode (TD1) is meant to be placed near the elbow, about 5 cm from the wrist, while the second (TD2) is positioned 10 cm further along the arm, near the wrist (Figure 6.3). The modulation of the temperature of individual actuators aims to elicit more complex illusory percepts between the two thermodes.

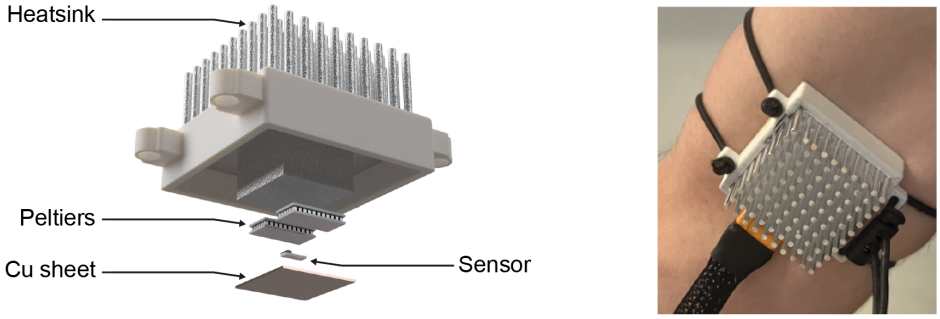


Figure 6.2: Prototype thermal display using thermoelectric modules and a passive heatsink

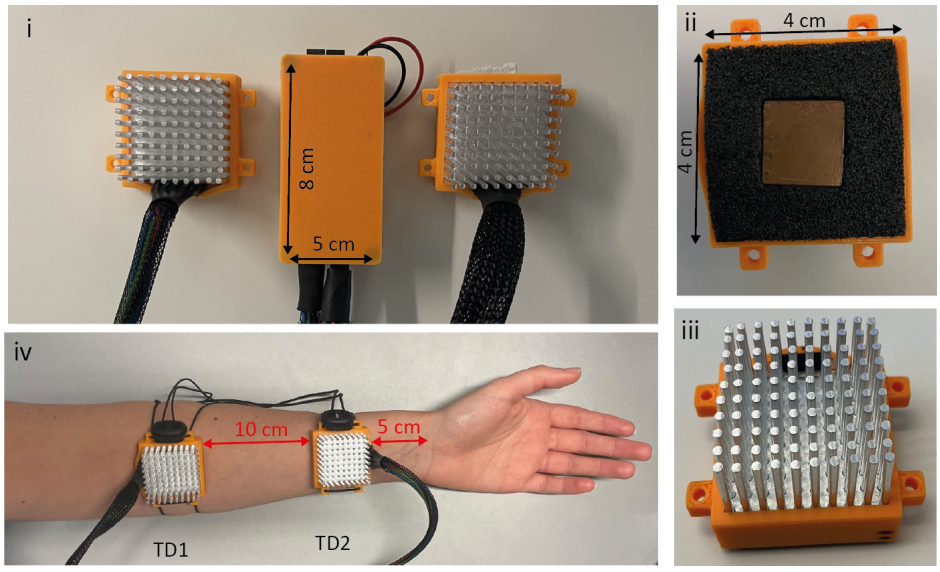


Figure 6.3: (i) The thermal feedback device, including the controller (orange box) and two thermodes. (ii) The Peltier element that modulates the skin temperature. (iii) A thermode equipped with a heatsink for effective heat dissipation. (iv) The thermal actuators are secured to the forearm using elastic bands.

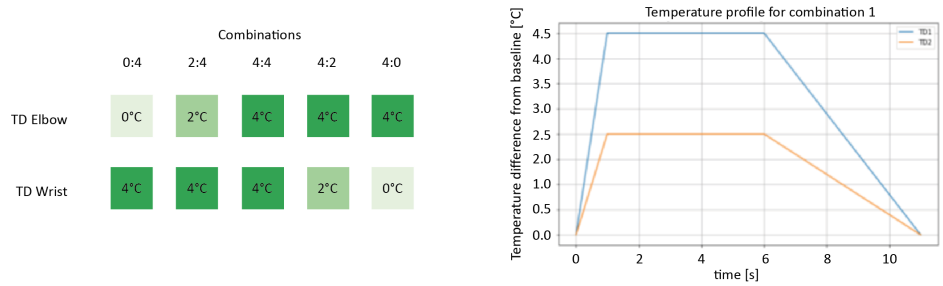


Figure 6.4: (Left) The different temperature combinations created for the experiment. The temperature (in °C) corresponds to the absolute difference from the participants' baseline skin temperature. (Right) Temperature profile for combination 4:2.

The concept of thermal illusions draws inspiration from well-established tactile illusions like the funneling illusion, which creates pseudo-tactile sensations by manipulating multiple stimuli on the body. While previous research has focused on tactile or combined thermo-tactile feedback, the effectiveness of thermal feedback alone in generating similar perceptual effects remains underexplored. The funneling technique adjusts the intensity and timing of stimuli to influence where the sensation is perceived, often toward the strongest stimulus. Here, we investigated the conditions under which such illusions could be elicited to provide more complex thermal sensations.

In this first experiment, the aim is to determine if varying thermal intensities between two thermodes can alter the perceived location of sensations. The hypothesis suggests that adjusting the intensity of thermal stimulation from two points can shift the perceived sensation to an intermediate area. Five combinations of thermal intensities (0°C, 2°C, and 4°C, relative to the participant's baseline skin temperature) were tested, with each combination applied for 11 seconds, accompanied by auditory cues to mark the temperature change (Figure 6.4).

Thermal sensitivity and the distribution of thermoreceptors vary across different parts of the body, particularly between proximal and distal areas of the forearm. To account for these differences, a calibration phase was included at the beginning of the experiments to determine the temperature differential (ΔT) between two thermodes, ensuring that participants perceived consistent temperatures at both the elbow and wrist. A staircase method was used for this calibration.

In the first experiment, 20 healthy participants (10 women and 10 men) aged 20 to 27 participated. Baseline skin temperatures at the wrist and elbow were measured using an infrared thermometer, and the thermodes were fixed to the non-dominant forearm. To avoid bias from visual cues, a box obscured the participant's forearm during

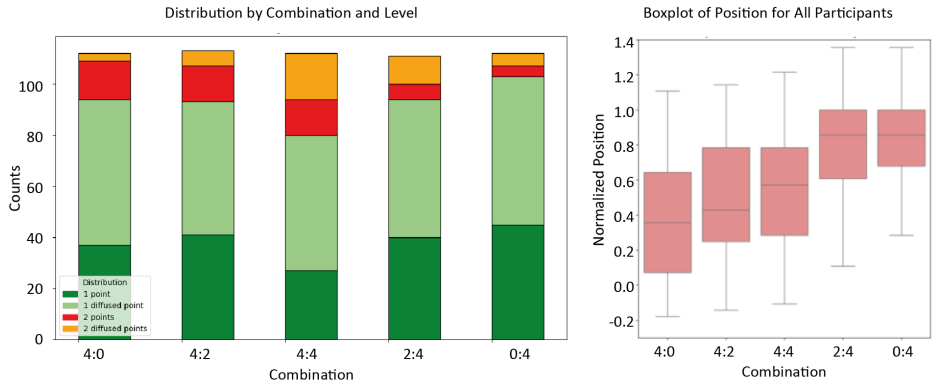


Figure 6.5: (Left) Participants' responses for the warm stimulations. The graph shows the stacked count of all participants' reported sensations for each combination. (Right) Boxplot of the normalized position felt by all participants for the different warm combinations.

the setup and experiment, with only the hand and elbow visible. A ruler was placed on top of the box to mark the wrist as the 0 cm reference point.

A Graphical User Interface (GUI) was developed to allow participants to immediately record their sensations following thermal stimulus exposure. Participants were instructed to use their right index finger to indicate, on the box, the perceived location of the temperature sensation, Additionally, participants reported the intensity and distribution of the sensation through the GUI, using forced-choice labels: 1 point, 1 diffused point, 2 points, or 2 diffused points.

6.2.1 Results

For warm stimulations, across all conditions, 82% of participants reported a single or diffused point of sensation, with 78% reporting a single-point perception when both thermodes were stimulated (Figure 6.5). Even in the most complex scenario (4:4 intensity), 68% reported one point or diffused sensation. To further analyze the illusion's effectiveness, participants' perceived positions of the sensations were mapped and normalized between the wrist and elbow. While a linear progression of sensations was observed, the study did not achieve clear differentiation across five distinct positions. Statistical tests (Wilcoxon Signed-Rank) revealed no significant differences between some combinations, indicating that while the intensity variations successfully induced intermediate sensations, the study could not generate five distinct perceptual points between the thermodes.

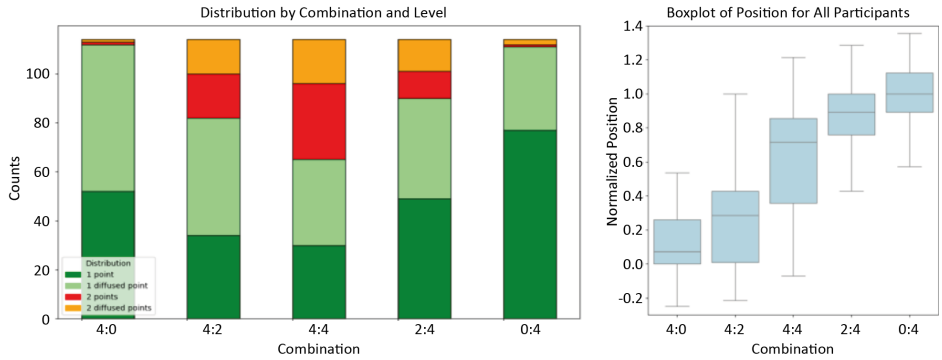


Figure 6.6: (Left) Participants' responses for the cold stimulations. The graph shows the stacked count of all participants' reported sensations for each combination. (Right) Boxplot of the normalized position felt by all participants for the different cold combinations.

As for the cold stimulations, across all conditions, 80% of sensations were reported as a single point or a diffused single point, consistent with expectations for conditions where only one thermode was stimulated (Figure 6.6). However, for combinations involving both thermodes, 68% reported a single point sensation, which is lower than the 78% observed in the warm modality. In particular, for combination 4:4 (both thermodes at -4°C), only 57% of participants reported a single or diffused single point. These results suggest that achieving an integrated single sensation is more challenging with cold stimulation than with warm. When comparing the positions of the perceived sensations, the data showed a good linear progression from wrist to elbow, indicating a clear distinction between different positions. Statistical analysis (Wilcoxon Signed-Rank tests) confirmed significant differences between the temperature combinations, successfully eliciting five distinct positions between the two thermodes. This demonstrates that varying the intensity of cold stimulation effectively induced intermediate sensations, albeit with lower integration success than in the warm modality.

6.3 Conclusions

In this chapter, we presented the design, development, and evaluation of a novel wearable thermal feedback system tailored for virtual reality applications. Through a combination of thermoelectric actuation, passive heat dissipation, and custom control electronics, our device delivers realistic warm and cold stimuli in a compact and portable form factor. The iterative prototyping process allowed us to refine both the hardware

and control strategies, resulting in a system that is not only wearable and energy-efficient but also capable of supporting complex perceptual experiments.

Our experimental investigation into thermal illusions demonstrated that varying the intensity of temperature stimuli between two spatially separated thermodes can modulate the perceived location of the sensation. While warm stimuli favored perceptual integration, cold stimuli provided clearer spatial differentiation—underscoring the potential of thermal feedback to create nuanced and dynamic somatosensory experiences in virtual environments.

Looking forward, this system offers a versatile platform for further exploration of temperature perception, including its integration with other modalities such as touch and force feedback. Its adaptability to different body locations and user sensitivities also makes it a promising tool for inclusive VR applications, particularly in rehabilitation and assistive technologies. In particular, regarding these scenarios, the presented technology has been implemented and validated in the Case Study "Extended Reality for People with Serious Mobility and Verbal Communication Diseases" (Chapter 22).

References

Cai, Shaoyu, Pingchuan Ke, Takuji Narumi, and Kening Zhu (2020). "Thermairglove: A pneumatic glove for thermal perception and material identification in virtual reality". In: *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 248–257.

Filingeri, Davide (July 2016). "Neurophysiology of skin thermal sensations". en. In: *Comprehensive Physiology* 6.3, pp. 1429–1491.

Gallo, Simon, Lucian Cucu, Nicolas Thevenaz, Ali Sengül, and Hannes Bleuler (2014). "Design and control of a novel thermo-tactile multimodal display". In: *2014 IEEE Haptics Symposium (Haptics)*, pp. 75–81.

Günther, Sebastian, Florian Müller, Dominik Schön, Omar Elmoghazy, Max Mühlhäuser, and Martin Schmitz (2020). "Therminator: Understanding the interdependency of visual and on-body thermal feedback in virtual reality". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Iberite, Francesco, Jonathan Muheim, Outman Akouissi, Simon Gallo, Giulio Rognini, Federico Morosato, André Clerc, Magnus Kalff, Emanuele Gruppioni, Silvestro Micera, and Solaiman Shokur (May 2023). "Restoration of natural thermal sensation in upper-limb amputees". en. In: *Science* 380.6646, pp. 731–735.

Lee, Yeonseo, Hyeonjung Lim, Yeunhee Kim, and Youngsu Cha (2020). "Thermal feedback system from robot hand for telepresence". In: *IEEE Access* 9, pp. 827–835.

Muheim, Jonathan, Francesco Iberite, Outman Akouissi, Rachel Monney, Federico Morosato, Emanuele Gruppioni, Silvestro Micera, and Solaiman Shokur (Feb. 2024). "A sensory-motor hand prosthesis with integrated thermal feedback". en. In: *Med* 5.2, 118–125.e5.

Peiris, Roshan Lalintha, Wei Peng, Zikun Chen, Liwei Chan, and Kouta Minamizawa (2017). "Thermovr: Exploring integrated thermal haptic feedback with head mounted displays". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5452–5456.

Ploumitsakou, Maria, Jonathan Muheim, Amanda Felouzis, Nerea-Isabel Carbonell Muñoz, Francesco Iberite, Outman Akouissi, Federico Morosato, Emanuele Gruppioni, Davide Filingeri, Silvestro Micera, and Solaiman Shokur (Mar. 2024). "Remapping wetness perception in upper limb amputees". en. In: *Advanced Intelligent Systems* 6.3, p. 2300512.

7. Wearable Haptics in Manipulation

Federica Serra[1], *Michael Efraimidis*[2], *Cristian Camardella*[1], *Katerina Mania*[2], *and Daniele Leonardis*[1]

[1] Scuola Superiore Sant'Anna (SSSA), Italy
[2] School of Electrical & Computer Engineering, Technical University of Crete (TUC), Greece

Abstract. Wearable haptic devices are essential for enhancing user immersion in virtual and augmented reality by delivering tactile cues that replicate or augment sensations experienced during physical interaction. These interfaces are commonly explored in Extended Reality systems to improve the realism of multisensory feedback. Their importance increases in applications involving fine motor skills, such as teleoperation, clinical simulation and training, and motor rehabilitation through serious games. A key challenge in designing such devices lies in providing rich tactile feedback without compromising user dexterity. This is due to the complexity of physical interactions, which involve a wide range of tactile cues and varying force dynamics, particularly during tasks like object manipulation. We propose two complementary wearable haptic solutions targeting the key sensations involved in object grasping: fine tactile perception of dynamic contact transients at the fingerpads, and kinesthetic force feedback during finger closure. For the former, we developed a lightweight, highly wearable haptic thimble that uses a flexible band actuated by direct-drive miniature DC motors, offering high dynamic response and low noise. For the latter, we present a soft actuated glove using compact nitinol tendon actuators to provide kinesthetic feedback at the level of finger movements.

7.1 Introduction

Hands and fingerpads are among the most sensitive areas of the human body and have been a major target for haptic interface design. The literature presents numerous solutions for fingertip devices capable of rendering vibration, contact force, surface texture, or directional feedback cues [Pacchierotti et al. 2017]. Achieving realistic interaction between a virtual object and the fingertip remains a technical challenge [Caldwell et al. 1997], primarily due to trade-offs among force fidelity, wearability, and compactness [Frisoli and Leonardis 2024].

Regarding force feedback gloves, the most commonly used solutions in the literature rely on electromagnetic DC (direct current) motors [Gu et al. 2016; Adilkhanov et al. 2022], which, however, result in relatively heavy, limiting the user's natural dexterity. Other solutions found also in commercially available force-feedback gloves (e.g., HaptX G1 and SenseGlove Nova 2) are based on passive brakes and tendons that typically constrain motion by opposing finger flexion when a virtual contact has to be rendered.

More experimental and advanced solutions are based on nitinol actuators, a nickel-titanium alloy that exhibits shape memory and superelastic properties. Nitinol components exert strong Newtonian forces at low weight and inertia, and has been successfully experimented as actuators in robotics [Engeberg et al. 2015] and manipulation [Zuo et al. 2024] scenarios. Nitinol-based haptic systems have been proposed: a 3 degrees-of-freedom (DOFs) wrist device uses nitinol actuators for force feedback in rehabilitation, yet it requires mineral oil cooling [Jeong et al. 2019]. Another produces finger skin pressure via ring elements for a single finger [Chernyshov et al. 2018], and in [Bourdot et al. 2018] a single-finger design emphasizes fast activation, although it has limits in scalability. While the thin and compact form factor shows high potential for wearable devices, challenges include complex shape training and the need for thermal management and regulation.

Regarding tactile feedback at fingerpads, different designs can be adopted, usually in the shape of thimble devices allowing more natural finger motion, and delivering skin-based cues without net kinesthetic force. Devices such as BeBop or Manus gloves offer lightweight vibrotactile feedback, although they are limited in rendering complex tactile dynamics.

To expand the range of sensations, several research prototypes have incorporated multiple DOFs, enabling directional skin stretch or contact reorientation through actively controlled platforms at the fingertip [Schorr et al. 2013; Gabardi et al. 2016]. While such devices enhance realism, they often suffer from increased bulk and limited wearability.

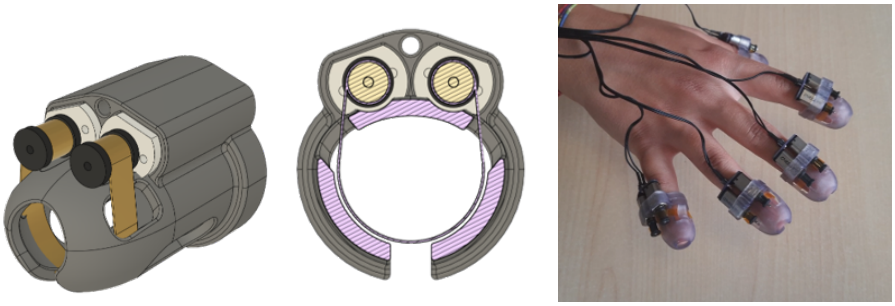
Multisensory commercial thimbles (e.g. Weart TouchDive, GoTouchVR) attempt to balance complexity and usability by combining contact, thermal, and vibration cues. Nonetheless, limitations remain in size, robustness, and long-term comfort. These challenges are particularly significant in rehabilitation, where patient dexterity may be impaired, and ease of use is paramount [Gutiérrez et al. 2021; Bortone et al. 2020]. Another relevant aspect is that the recent advent of embedded hand-tracking systems in Virtual reality (VR) headsets (e.g., Microsoft HoloLens, Meta Oculus) allows natural manipulation without external markers, fostering robust and rapid usability of these systems. For this reason, newer haptic designs have to prioritize minimal changes to the hand shape in order to support compatibility with vision-based tracking systems.

7.2 Methodology

Two complementary devices have been developed to cover the main haptic perception during object grasping and manipulation. The first is a lightweight and compact thimble, with design aimed at enhancing the quality and linearity of the feedback, rather than on the absolute intensity of the forces. Within project SUN, such technology was developed in particular to provide tactile feedback in virtual manipulation exercises for rehabilitation (Case Study "Extended Reality for Rehabilitation" described in Chapter 20). The second device is a haptic glove designed to deliver kinesthetic force feedback at the level of finger movement. It implements nitinol tendons, leveraging the compactness and flexibility of such an actuation solution.

7.2.1 Direct-drive Haptic Thimble

The component is designed as a soft, thimble-shaped unit, one for each finger of the hand, with built-in miniaturized actuators (Figure 7.1). Actuators are two twin rotary electromagnetic motors, operating at low voltage to be compliant with battery operation as a wearable device. The output shaft of each motor is connected to a flexible band that wraps around the fingertip; the section view depicting such a mechanism is shown in Figure 7.1b. The combined drive of the two motors in opposite directions of rotation causes the soft band to squeeze the finger. Then, the pressure of the band can be modulated by the intensity of the current flowing through the motor coils. SLA 3D printing with soft resin was used to manufacture the thimble. The high level of detail of the printing method allowed for a more structured shape of the thimble: the flexible side lobes serve to adapt the size to different finger diameters. In addition, the cavities and passages in the structure aim to improve fit (the flexible band retains its shape when the thimble is empty) and drive efficiency, keeping the flexible band free



(a) (b) (c)

Figure 7.1: (a) The developed Actuated Band Thimble with soft actuated band. (b) section view of the soft band and direct-drive mechanism, (c) The implemented Actuated Band Thimble covering five fingers.

to slide within the thimble structure. Different sizes of thimbles have been produced with a connector that simplifies the change of thimbles for different users. Figure 7.1c shows the prototyping of the component in five thimbles. The electronics are compact and can be worn at the level of the forearm.

7.2.2 Wireless Haptic Glove using Nitinol as a Force Feedback Actuator

The developed device consists of a lightweight five-finger tactile glove that uses nitinol actuators for force feedback through fingerprints (Figure 7.2). The glove mimics the musculoskeletal architecture through thermally activated nitinol springs mounted on the forearm. The main contributions are:

- A nitinol-actuated glove that provides force feedback to the five fingers via the fingertips, integrated with a head-mounted augmented reality (AR) display (HMD);
- A 3D-printed porous thermoplastic elastomer (TPE) bellows that provides cooling, flexibility, and electrical insulation;
- Wireless, battery-powered, allowing for mobile and immersive use;
- A psychophysical evaluation using grasping and piano tasks with real-time temperature monitoring;
- A 280-gram glove that improves usability compared to heavier motor-based models.

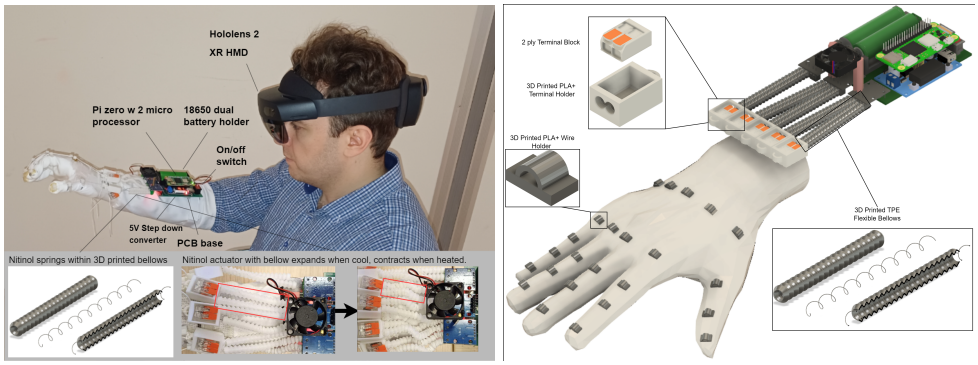


Figure 7.2: (a) 3D view of the haptic glove. The 3D printed parts are printed using solid PLA+ and flexible TPE filaments. (b) Prototype of the haptic glove making use of the nitinol-based actuator.

A lightweight haptic glove was developed using nitinol springs to deliver force feedback by contracting between static and moving finger platforms without impeding motion. A Raspberry Pi Zero W 2 controls heating, powered by a dual L9110S motor driver (800mA/channel). Two Murata US18650VTC6 batteries in series provide 7.4 V and up to 6 A. A custom PCB routes high current using 3mm-wide traces and includes a 5 V converter for the Pi. Cooling is supported by a 5 V, 0.2 A fan and porous 3D-printed bellows.

Cooling and Heating Methodology Passive cooling delayed reset times, taking up to 10 s for nitinol to return to its martensitic state. Reducing wire diameter, improves response. The 0.1 mm wire would break with minimal finger force. A 0.25 mm, 55°C wire retained shape and provided durable 3N feedback in a 1-second activation time. A 5 V fan reduced the cooling time to 2 seconds. Single/dual channel current outputs were 0.8 A and 1.23 A, respectively. For tests, a single channel was used. 0.75 mm, 0.5 mm, and 0.25 mm wires retracted in 11s, 4s, and 1 s, respectively; 0.1 mm retracted in 0.1 s; however, it would break with minimum finger force. For that purpose, the next wire candidate was used, which is the 0.25 mm wire at 55°C. With a fan, the 0.25 mm cooled in 2 s, which was optimal in strength and cooling rate. Joule heating follows $H = I^2Rt$ where $R = \frac{\rho L}{A}$ and $A = \pi(d/2)^2$ [Wilson and Hernández-Hall 2014], so heat increases with length and decreases with d^2.

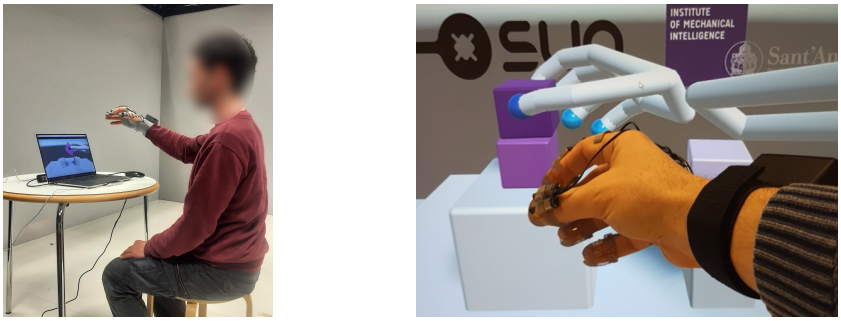
Training Nitinol Training nitinol required heating to 600 °C for 3 min using a 50 Hz 1800 W RS 18000D heat gun, with the wire wrapped on a 4 mm screw with 1.5 mm pitch. A fire brick served as the base for securing the screws during heating, and a Uni-T721M smartphone thermal camera confirmed a surface temperature of 320 °C When the surface temperature was reached in approximately 3 minutes, water was used to instantly cool down the material. Trained springs retained their form without brittleness. Finger actuators measured 10–11 cm, with 6–10 coils (thumb: 6, middle: 10). Coil spacing was 1.1 mm (thumb: 1.6 mm), with 4 mm diameters. Stiffness is given by $k = \frac{Gd^4}{8N_aD^3}$ [Nazir et al. 2020], so more coils allow longer stroke. TPE bellows were 3D-printed to isolate each spring to prevent short circuits and shape reprogramming.

Software Setup HoloLens 2's hand tracking was used. Early prototypes placed nitinol components on the glove but occluded hand tracking. Moving them to the forearm resolved this. Interactive 3D scenes were developed in Unity 3D via Microsoft MRTK 3 to evaluate force feedback perception via psychophysics. The system uses bidirectional UDP communication between the Pi and HoloLens 2. Scenes requiring force feedback were presented: one where users grasped and released a virtual milk carton, and another where they played piano.

7.3 Direct-drive Haptic Thimble Experiments

The effectiveness of the haptic thimbles was evaluated through a fine manipulation virtual task based on a pick-and-place exercise with full physical simulation, requiring higher precision than facilitated paradigms and allowing more realistic force interactions to be conveyed to the user. In the setup (Figure 7.3a), participants wore the haptic device and interacted with a virtual cube (50 mm edge) to be moved onto a platform 200 mm away from the starting position (Figure 7.3b). Hand tracking was performed using LeapMotion, with contact limited to thumb and index fingers. Interaction forces were computed via virtual spring coupling and sent to the haptic devices at 100 Hz via Wi-Fi.

Four participants (2 males, 2 females, aged 24–39) completed 15 trials under two conditions: HV- haptic + visual feedback and V- visual feedback only. Each trial could result in: Correct - cube correctly placed; Fallen - cube missed the platform; Broken cube - the indentation of the virtual fingers (pinch distance controlled by the user) was higher than a given threshold (5 mm) for more than 1 s.



(a) (b)

Figure 7.3: (a) Experimental setup, and (b) screenshot of the virtual pick and place task implementing physical simulation and modulated tactile feedback related to exerted finger forces

7.3.1 Results and Discussions

As regards the experimental evaluation of the Direct-drive Haptic Thimble, results are reported in Figure 7.4a. It emerges that the correct ratio of the task was higher in the presence of the haptic feedback, with respect to the visual-only condition. Moreover, it is interesting to note how the distribution of the incorrect repetitions (between broken or fallen cubes) changes in the presence of the haptic feedback. Without feedback, participants tended to exceed with the virtual pinching forces, probably aiming at a more secure grasping. This led to a higher number of broken cubes with respect to the fallen ones. With Haptic feedback enabled, both the limits imposed by the setup (pinch forces not too loose or not too high) led to more balanced results. The above results are confirmed by the measurements of the virtual interaction forces. These are significantly reduced in the presence of the haptic feedback (paired T-Test, Force Mean HV = 5.2 mm, V = 7.1 mm, $p<0.05$). It suggests that the participants were provided a coherent information by the haptic feedback, mainly related to the contact threshold and to the modulation of the grasping force (Figure 7.4b). It is also worth noting that this trend aligns with a more natural execution of the grasping task. Studies have shown that, during natural grasping, the applied forces are maintained just above the minimum level of static friction needed to prevent slippage [Cole and Abbs 1988]. However, in situations where sensory feedback is limited, such as in virtual environments, the safety margin increases, resulting in higher forces being applied to securely grasp and lift the object. In addition, the presence of the additional information provided by haptic feedback resulted in a more repeatable grasping action performed by the participants,

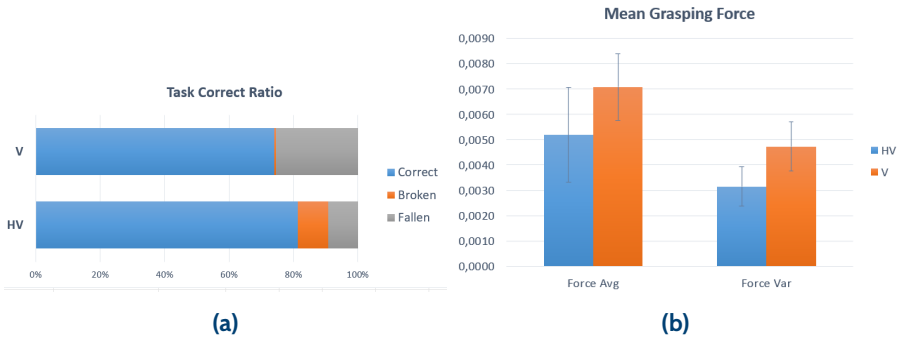


Figure 7.4: Results of the pick-and-place task: (a) Correct rate measured in the pick-and-place tasks. (b) The mean values of the virtual grasping force. With the haptic feedback enabled (HV) interaction forces are reduced.

with a significantly reduced variation of the virtual grasping force measured between repetitions (paired T-Test, Force Var HV = 3.1 mm, V = 4.7 mm, $p<0.01$).

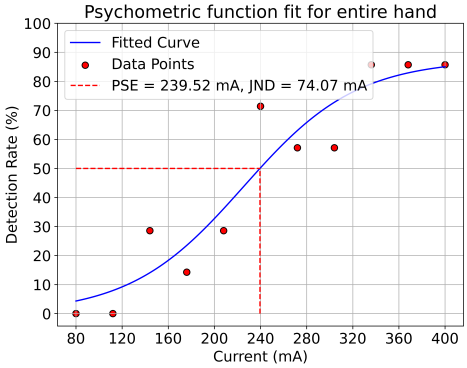
7.4 Nitinol-Based Haptic Glove Experiments

The nitinol-based haptic glove was evaluated using psychophysics, which studies the relationship between stimulus intensity and human perception. A key metric is the Just Noticeable Difference (JND), the smallest detectable change in stimulus. Since human perception is variable, JND is defined statistically. We employed the method of constant stimuli, presenting randomized currents of varying intensity to users. Detection probability was plotted against stimulus intensity to fit a psychometric (Gaussian) curve. The JND was defined as the difference in current between 50% and 75% detection thresholds.

7.4.1 Hardware Setup and Electronic Design

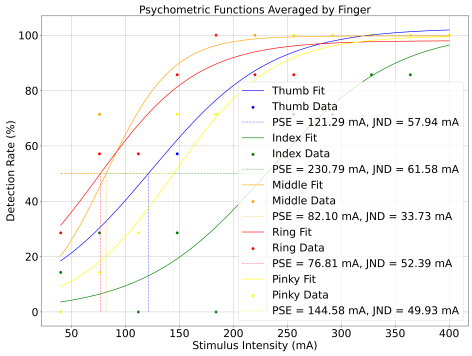
Experiment 1: Grab and Release *Setup:* 7 right-handed participants (6 male, 1 female; mean age 28) wore a HoloLens 2 and the haptic glove. After calibrating the cloth wires to fit each hand, users grabbed a digital milk carton and placed it on a virtual shelf (Figure 7.5a). For each trial, they indicated whether they felt force feedback. Stimuli ranged from 40 mA to 400 mA in 40 mA steps (11 levels), presented in random order, repeated 7 times per participant. Zero current trials were excluded. A singular threshold was computed for each user using their psychometric function (Figure 7.5b).

(a) The grab and release experiment setup



(b) Psychometric function fit for one participant (P5) showing a JND of 74.07 mA

(c) The piano-playing experiment setup



(d) Psychometric function fit for one participant (P6). The observed JNDs are 57.94 mA (Thumb), 61.58 mA (Index), 33.73 mA (Middle), 52.39 mA (Ring), and 49.93 mA (Pinky)

Figure 7.5: Overview of our experimental setup and results.

Experiment 2: Piano Playing *Setup:* 7 participants (3 male, 4 female; mean age 30) used the glove to play virtual piano keys (Figure 7.5c). Currents (40–400 mA) were applied to individual fingers in random order, repeated 7 times. Each finger was calibrated beforehand. When HoloLens 2's ring and pinky finger showed poor tracking, other actuators were disabled to isolate feedback.

7.4.2 Results and Discussions

Experiment 1: Grab and Release *Results:* The mean JND was 20.80 mA (SD = 10.72 mA). Each participant completed 77 trials (539 total). Currents below this threshold were not reliably perceived when all actuators were active.

Experiment 2: Piano Playing *Results:* The average JND values were 50.48 mA (SD = 48.05) for the thumb, 83.92 mA (SD = 30.01) for the index, 58.71 mA (SD = 64.78) for the middle, 43.89 mA (SD = 25.40) for the ring and 52.03 mA (SD = 21.70) for the pinky. The pinky was most sensitive (lowest JND), and the middle finger the least. Users reported varying force intensities with current changes, indicating a capacity for dynamic force feedback perception (Figure 7.5d).

Compared to existing haptic solutions in the literature, such as Springlets [Hamdan et al. 2019], which suffers from slow activation and cooldown cycles, or more complex powering and cooling methods [Jeong et al. 2019], the developed glove offers a lighter, safer, and faster alternative. The proposed device is fully wireless, battery-powered, and actively cooled via a compact air-based system, providing 3N of force feedback per finger in a 280g form factor. The implemented flexible 3D-printed bellows protect components and the user's skin. Regarding usability, participants appreciated the glove's low weight and visible actuation but noted tracking issues on the pinky and ring fingers due to HoloLens's limits. Psychophysical tests confirmed perception of both full-hand and individual finger feedback. Though designed for low-dynamics feedback, users reported variable force sensations, suggesting dynamic potential with added thermal and force-resistive sensors. Future improvements include dorsal actuator placement for improved motion control and Peltier modules or PEG-based water solution offering significantly faster cooling.

7.5 Conclusions

Haptic feedback is a complex sensory channel as regards the variety of physical interactions occurring at our hands and generating different and informative sensations. Here we presented two complementary wearable haptic devices addressing two main aspects of tactile feedback in manipulation: the perception and modulation of the grasping force, addressed by an actuated glove and providing force-feedback at the level of the whole kinematic chain of the fingers, and the perception of fine contact transients occurring at the fingerpads, addressed by compact and wide-bandwidth haptic thimbles. Both devices implement specific design solutions aimed at increasing wearability, such as the nitinol actuators for the glove, and the direct-drive, flexible band for the thimble. It results in a trade-off between rendering capabilities, such as intensity, bandwidth, number of degrees of freedom, and compactness and usability, which remains the main challenge in the field. Nonetheless, experiments with the two novel devices proved the effectiveness of the rendered feedback given the compact and lightweight shape of the device. The glove enabled AR interaction with kinesthetic feedback in sample AR scenarios, and has potential for further use in Extended Reality (XR) rehabilitation and soft robotics. The proposed haptic thimble as well shown significant improvements in performance metrics measured in a virtual fine pick and place task, suggesting the feedback was providing additional, meaningful contact information to the participant without degrading dexterity. Within the project SUN, the haptic thimble has been implemented and validated in the Case Study "Extended Reality for Rehabilitation" (Chapter 20).

References

Adilkhanov, Adilzhan, Matteo Rubagotti, and Zhanat Kappassov (2022). "Haptic Devices: Wearability-Based Taxonomy and Literature Review". en. In: *IEEE Access* 10, pp. 91923–91947.

Bortone, Ilaria, Michele Barsotti, Daniele Leonardis, Alessandra Crecchi, Alessandra Tozzini, Luca Bonfiglio, and Antonio Frisoli (2020). "Immersive Virtual Environments and Wearable Haptic Devices in rehabilitation of children with neuromotor impairments: a single-blind randomized controlled crossover pilot study". In: *Journal of NeuroEngineering and Rehabilitation* 17.1, pp. 1–14.

Bourdot, Patrick, Sue Cobb, Victoria Interrante, Didier Stricker, et al. (2018). *Virtual Reality and Augmented Reality: 15th EuroVR International Conference, EuroVR 2018, London, UK, October 22–23, 2018, Proceedings*. Vol. 11162. Springer.

Caldwell, Darwin G, N Tsagarakis, and Andrew Wardle (1997). "Mechano thermo and proprioceptor feedback for integrated haptic feedback". In: *Proceedings of International Conference on Robotics and Automation*. Vol. 3. IEEE, pp. 2491–2496.

Chernyshov, George, Benjamin Tag, Cedric Caremel, Feier Cao, Gemma Liu, and Kai Kunze (Oct. 2018). "Shape memory alloy wire actuators for soft, wearable haptic devices". en. In: *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. Singapore Singapore: ACM, pp. 112–119.

Cole, Kelly J and James H Abbs (1988). "Grip force adjustments evoked by load force perturbations of a grasped object". In: *Journal of neurophysiology* 60.4, pp. 1513–1522.

Engeberg, Erik D, Savas Dilibal, Morteza Vatani, Jae-Won Choi, and John Lavery (Aug. 2015). "Anthropomorphic finger antagonistically actuated by SMA plates". In: *Bioinspiration & Biomimetics* 10.5, p. 056002.

Frisoli, Antonio and Daniele Leonardis (2024). "Wearable haptics for virtual reality and beyond". In: *Nature Reviews Electrical Engineering* 1.10, pp. 666–679.

Gabardi, Massimiliano, Massimiliano Solazzi, Daniele Leonardis, and Antonio Frisoli (2016). "A new wearable fingertip haptic interface for the rendering of virtual shapes and surface features". In: *Haptics Symposium (HAPTICS), 2016 IEEE*. IEEE, pp. 140–146.

Gu, Xiaochi, Yifei Zhang, Weize Sun, Yuanzhe Bian, Dao Zhou, and Per Ola Kristensson (May 2016). "Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR". In: San Jose California USA: ACM, pp. 1991–1995.

Gutiérrez, Álvaro, Nicola Farella, Ángel Gil-Agudo, and Ana de los Reyes Guzmán (2021). "Virtual Reality Environment with Haptic Feedback Thimble for Post Spinal Cord Injury Upper-Limb Rehabilitation". In: *Applied Sciences* 11.6, p. 2476.

Hamdan, Nur Al-huda, Adrian Wagner, Simon Voelker, Jürgen Steimle, and Jan Borchers (May 2019). "Springlets: Expressive, Flexible and Silent On-Skin Tactile Interfaces". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, pp. 1–14.

Jeong, Jaeyeon, Ibrahim Bin Yasir, Jungwoo Han, Cheol Hoon Park, Soo-Kyung Bok, and Ki-Uk Kyung (Sept. 2019). "Design of Shape Memory Alloy-Based Soft Wearable Robot for Assisting Wrist Motion". In: *Applied Sciences* 9.19, p. 4025.

Nazir, Aamer, Mubasher Ali, Chih-Hua Hsieh, and Jeng-Ywan Jeng (Oct. 2020). "Investigation of stiffness and energy absorption of variable dimension helical springs fabricated using multijet fusion technology". In: *The International Journal of Advanced Manufacturing Technology* 110.9–10, pp. 2591–2602.

Pacchierotti, Claudio, Stephen Sinclair, Massimiliano Solazzi, Antonio Frisoli, Vincent Hayward, and Domenico Prattichizzo (2017). "Wearable haptic systems for the fin-

gertip and the hand: taxonomy, review, and perspectives". In: *IEEE transactions on haptics* 10.4, pp. 580–600.

Schorr, Samuel B, Zhan Fan Quek, Robert Y Romano, Ilana Nisky, William R Provancher, and Allison M Okamura (2013). "Sensory substitution via cutaneous skin stretch feedback". In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, pp. 2341–2346.

Wilson, J.D. and C.A. Hernández-Hall (2014). *Physics Laboratory Experiments*. Cengage Learning.

Zuo, Zonghao, Xia He, Haoxuan Wang, Zhuyin Shao, Jiaqi Liu, Qiyi Zhang, Fei Pan, and Li Wen (Mar. 2024). "A Nitinol-Embedded Wearable Soft Robotic Gripper for Deep-Sea Manipulation: A Wearable Device for Deep-Sea Delicate Operation". In: *IEEE Robotics & Automation Magazine* 31.1, pp. 96–107.

8. Distributed Wearable Haptics

Federica Serra[1], *Ali KhalilianMotamed Bonab*[1], *Cristian Camardella*[1], *and Daniele Leonardis*[1]

[1] Scuola Superiore Sant'Anna (SSSA), Italy

Abstract. While the conventional role of a haptic device is to simulate the sense of touch perceived during physical interaction, the tactile sensory pathways can be used as well to convey high-level information related, for example, to situational awareness, body posture, or navigation. We present here two armband haptic devices based on complementary rendering approaches, and designed to convey directional and symbolic information to the user, particularly for Extended Reality (XR) and rehabilitation motor tasks. The wrist-worn armband integrates four micro servomotors and flexible bands to provide shallow torques at the ulnar-radial and flexion-extension wrist movements. Experiments conducted in a Virtual Reality (VR) scenario with healthy participants showed that haptic feedback significantly enhanced movement accuracy and reduced variability compared to no-feedback conditions, with performance comparable to visual guidance. The second device is based on two direct-drive motors and a flexible band wrapped around the forearm, capable of delivering low-noise, continuous to fast dynamic haptic stimuli. Tests in a VR rehabilitation exercise with healthy subjects during a grasp-and-pour task confirmed that tactile feedback, particularly the combined mode, improved accuracy and reduced inter-subject variability. Overall, both devices demonstrate the effectiveness of distributed tactile guidance in immersive VR environments. These results highlight the potential of wearable haptics to support motor control and improve rehabilitation outcomes, offering a compelling tool for enhancing user engagement and interaction in VR-based therapeutic applications.

8.1 Introduction

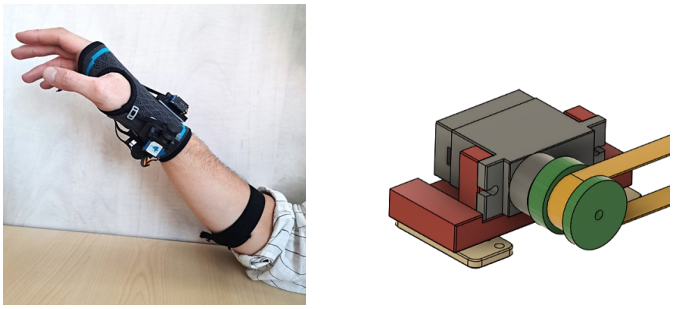
Distributed wearable haptics aim at delivering global, high-level cues such as motion direction, symbolic feedback, or guidance, by combining stimuli delivered at different locations of the body. These devices preserve hand mobility and are particularly suitable for VR-based training and rehabilitation, where they can provide postural feedback to the user. In project SUN, such technologies have been validated in the Case Studies "Extended Reality for Rehabilitation" (Chapter 20) and "Extended Reality for People with Serious Mobility and Verbal Communication Diseases (Chapter 22).

Recent developments have explored the use of wearable haptic devices not only for navigation and interaction in virtual environments, but also for motor guidance in rehabilitation contexts [Eguchi et al. 2023; Scheggi et al. 2014; Bortone et al. 2020]. These systems provide tactile signals to promote correct movement execution, compensating for the absence of direct supervision by the therapist in home settings [Chen et al. 2019; Levin et al. 2015]. In particular, in immersive virtual reality, where the visual channel is already heavily stimulated, tactile feedback represents an intuitive and underutilised alternative for movement guidance [Frisoli and Leonardis 2024; Vitense et al. 2003].

Thus, a central objective in the development of next-generation wearable haptics is to optimize the trade-off between the richness of haptic feedback and minimal encumbrance and mass, preserving both dexterity and comfort of the user. To convey perceivable, localized tactile stimuli, wearable devices rely on skin-stretch mechanisms, such as motor-driven belts [Casini et al. 2015; Aggravi et al. 2018; Meli et al. 2018], or mechanical linkages with multi-bar systems [Moriyama et al. 2018] or rotating tactors [Chinello et al. 2017] to convey directional information. More innovative designs include mobile units like Movelet, a wheeled bracelet producing local skin stretch [Dobbelstein et al. 2018], or piezo-driven rotational tactors for compact feedback [Bark et al. 2010]. Handheld haptic tools can also transmit cues via direct forces on the hand or wrist [Spiers and Dollar 2016; Koslover et al. 2011; Provancher 2014].

8.2 Methodology and Results

We propose here two novel devices, a wristband and an armband device, implementing actuated belts with two different rendering approaches, with the aim of providing directional cues by combining distributed actuation points at the forearm.



(a) (b)

Figure 8.1: (a) Configuration of four devices to provide directional hints to the wrist in two directions. One pair of haptic modules is used to provide directional signals related to the ulnar-radial deviation of the wrist, while another pair provides signals related to flexion-extension. (b) Single actuator module.

8.2.1 Haptic Wristband Device for Shallow Torque Rendering

The first component is a wearable electromechanical device designed for generating tactile sensations at the wrist through the shallow torques approach, with the aim of providing directional and intensity-based tactile information. It is intended for motor rehabilitation in virtual environments, typically in rehabilitative serious games, where the feedback can be directly related to the proposed motor exercise.

The developed device consists of an elastic glove to which four haptic modules are fixed to provide rotational feedback at the wrist. Each haptic module comprises a micro servomotor with a pulley around which a non-elastic band is wrapped, and a flexible plate that secures the haptic module to the glove. To facilitate the wearing and removal of the glove on the hands of different sizes, a sliding mechanism has been designed to facilitate the insertion and removal of the motor from its housing. The control electronics and power supply battery are secured to the arm with an elastic strap. The inertial measurement unit (IMU) sensor is positioned on the palm of the hand to measure the wrist's orientation in real time (Figure 8.1).

The system is designed to optimize wearability, weight, and balancing by distributing the bulk and mass of the components around the wrist, and by relocating the battery and wireless control electronics to the arm.

The flexible plate and buckles are made using 3D printing with Filaflex TPU 82A polymer material, while the pulley and the sliding mechanism are 3D printed in Onyx. The non-elastic bands pass through two buckles anchored to the glove; once the motor is activated, the pulley rotates, putting tension on the band. This allows the system of

four actuators to apply shallow torque to the anatomical wrist articulation, providing clear rotational hints. The four actuators are miniature commercial RC servo controls made of plastic and metal (Model MG92B distributed by Adafruit Industries Inc.). Each servo features a DC electromagnetic rotary motor, a gear reduction, a potentiometer for measuring the angular position of the output shaft, and an integrated control electronics board.

Experimental Results of the Wristband Haptic Device

To evaluate the effectiveness of the method, an experiment was conducted in a virtual reality scenario (Figure 8.2). The experiment took place on 10 healthy volunteers (two women and eight men), aged between 28 and 38 years. Each subject was asked to sit in front of a desk and a computer screen, which was used to provide visual signals and display the virtual environment. Subsequently, the subject had to wear the previously described haptic system, the control electronics on their arm, and the IMU sensor on the palm of their hand.

The experiment consisted of trials during which the subject had to guide an airplane within a virtual scenario, moving their wrist to try to reach five specific targets in space. At each sample time, the subject received directional feedback on their wrist, given by the angular difference between the pointing vector of the airplane and the vector linking the center of the airplane with the target. In this way, the subject continuously received cues on the direction of rotation and the amplitude associated with the two degrees of freedom (DOFs) considered (ulnar-radial deviation and flexion-extension). Three experimental conditions were proposed: only haptic feedback (HH condition), without visual/haptic feedback (NH condition), and only visual feedback (VH condition, with visual arrows on the screen).

Data acquired during the experimental session were processed to extract relevant features. In particular, the effectiveness of directional feedback received by the subject was

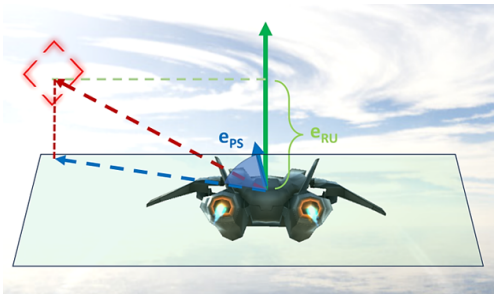


Figure 8.2: The developed virtual reality scenario

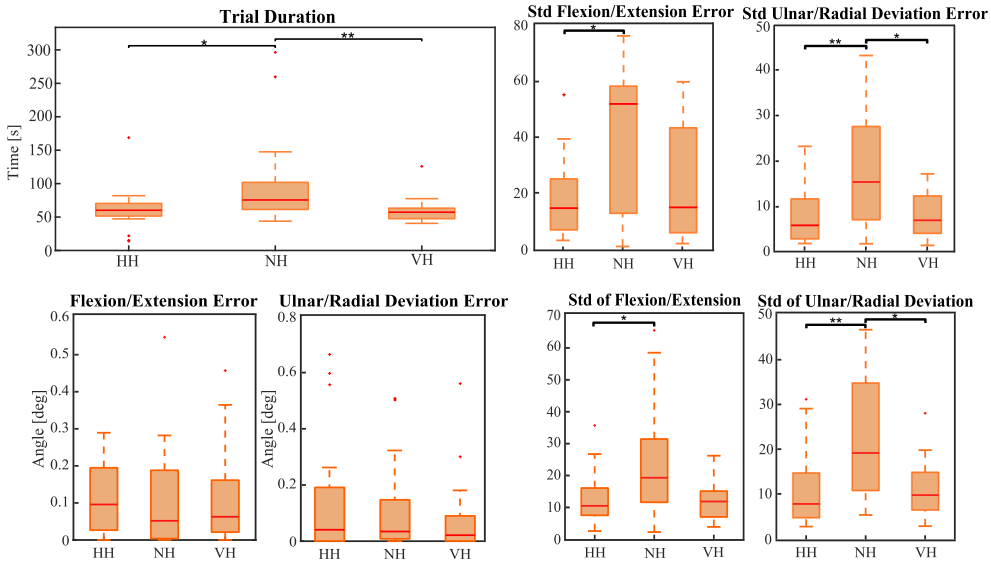


Figure 8.3: Performance metrics graphs across the three experimental conditions

evaluated by comparing wrist angular differences in the haptic-only versus visual-only condition and haptic-only versus without visual-haptic condition. Only wrist angular differences on the 2 DOFs associated with ulno-radial deviation DOF and flexion-extension DOF were considered. In addition, extracted features consist of: mean and standard deviation of the angular difference on each of the 2-DOF considered, the time required to reach the virtual targets during each trial, and variability of the wrist rotation during each target-reaching task.

The results obtained from the three different conditions tested indicate that, under the HH condition, there are significant improvements compared to the NH condition in terms of time duration, standard deviation of wrist rotation, and variability of wrist rotation. Additionally, the results under the HH condition are comparable to those obtained under the VH condition. This suggests that the haptic device has the potential to guide wrist movements with an effectiveness comparable to the visual feedback provided in the VH condition (Figure 8.3).

8.2.2 Haptic Armband for Directional Feedback

The developed armband haptic device is designed to generate tactile sensations at the arm or forearm segment (Figure 8.4a) by actuating a flexible belt around the limb. This haptic device features a bracelet structure that houses actuators positioned around the

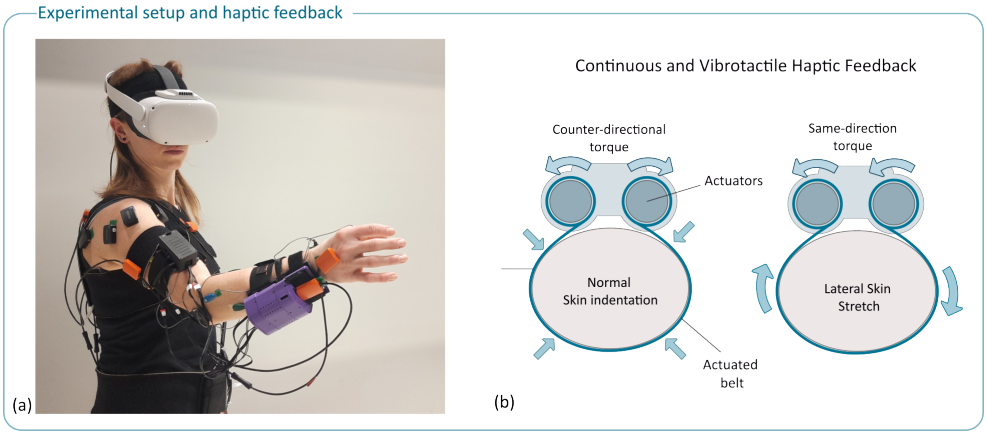


Figure 8.4: (a) Armband based on vibrotactile feedback. (b) Operation of the wearable haptic device, which provides squeezing and lateral stretching stimuli.

user's forearm. The actuators consist of rotary electromagnetic motors fixed to the structure of the bracelet, which is secured to the user's forearm using velcro straps. Compared to similar designs presented in literature [Casini et al. 2015], the presented approach features direct-drive motors in place of gearmotors, enabling dynamic haptic stimuli, including vibrations.

The movable part of each actuator is connected to the user's arm through a flexible element, which is pulled by the actuator itself to transmit tactile stimuli, such as transients and vibrations. The combined activation of multiple actuators also aims to send stimulation patterns able to convey various types of information to the user, such as directions or different symbolic information.

The system utilizes a flexible structure made of elastic bands and rigid elements that support actuators and control electronics. The rigid elements are produced using 3D printing with nylon filament (Onyx material produced by Markforged). The actuators are commercial DC electric motors operating in the very low voltage range (3.7 V DC), manufactured by Maxon Motors Group (Model DCX22S, nominal torque 14.4 mNm).

Experimental Results - Armband Device

We evaluated the effects of haptic guidance in a pick-and-place exercise with prono-supination in a group of ten healthy subjects. The exercise was conducted in VR within a serious game designed for rehabilitation. An analysis was conducted in which objective metrics of performance were measured, assessing the ability of feedback to effectively guide posture. This evaluation was conducted across five conditions,

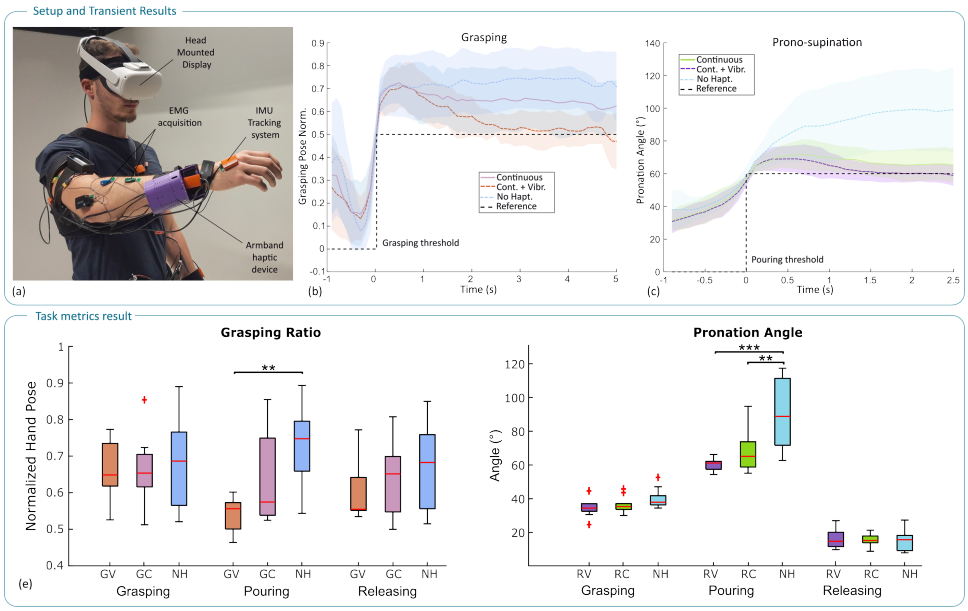


Figure 8.5: (a) The experimental setup includes monitoring of upper limb kinematics and muscle activation via sEMG, as well as visual tracking via a 3D head-mounted display. (b) The transient signals mediated by the initial grasping phase show how the hand adapts to the different feedback provided to achieve the target pose. (c) The pronation-supination signals at the beginning of the pouring phase highlight the guiding effect of feedback. (e) Aggregate results related to grip modulation and pronation-supination.

investigating guidance on two motor actions (grasping and prono-supination) and in two different feedback modalities, continuous (C) and combined continuous plus vibrotactile feedback (V). The five conditions were: grasping continuous (GC), grasping continuous + vibrotactile feedback (GV), no haptic feedback (NH), rolling continuous (RC), and rolling continuous + vibrotactile feedback (RV) (Figure 8.5).

With the experimental setup depicted in Figure 8.5a and throughout each exercise phase, we monitored two key metrics: the hand grasping ratio, normalized from 0 (fully open) to 1 (fully closed), and the hand prono-supination angle (positive angles toward pronation). The sequences of activities in virtual reality are: (1) grasp the ingredient, (2) move over the cauldron, (3) rotate to pour, (4) return to the starting position, (5) pause for 5 seconds. Transients depicted in Figure 8.5b and Figure 8.5c clearly show the reaction of the subject to the onset of the feedback at the beginning of the grasping phase (phase 1 to phase 2) and of the prono-supination phase (phase 2 to phase 3). In all the feedback conditions, after an initial overshoot, subjects adjusted the hand pose closer to the reference. The continuous vibrotactile feedback appears the most effective

as regards precision, with the pose error reduced below 10% in $2.5\ s$ and $1.2\ s$ for the grasping and prono-supination modulation, respectively. Notably, the high within-subject variability observed under the NH condition (green area in Figure 8.5b and c) further underscores the importance of haptic motor guidance in achieving optimal performance. Results aggregated for each phase and condition, depicted in Figure 8.5e confirm the transient results. Starting from phase 2 and throughout phase 3 and phase 4, participants modulated the grasping pose effectively in particular with the V feedback, achieving hand poses that were closer to the desired reference and with noticeably smaller variance between subjects; the V feedback resulted significantly closer to the reference in phase 3 (phase 2 - GV (0.66 $\pm$ 0.08), GC (0.66 $\pm$ 0.10), NF (0.69 $\pm$ 0.13); phase 3 - GV (0.53 $\pm$ 0.07), GC (0.63 $\pm$ 0.12), NF (0.73 $\pm$ 0.12), $p = 0.002$ for GV vs NF; phase 4 - GV (0.59 $\pm$ 0.08), GC (0.64 $\pm$ 0.11), NF (0.67 $\pm$ 0.12)).

Haptic feedback played a crucial role in guiding participants to achieve the precise pouring angle associated with proper prono-supination in phases 2 and 3. As shown in Figure 8.5e, the mean pouring angle was significantly closer to the reference in the RV condition and in the RC condition, compared to NH (phase 2 - RV (36.25 $\pm$ 6.13), RC (37.34 $\pm$ 4.89), NF (40.80 $\pm$ 5.80); phase 3 - RV (61.60 $\pm$ 3.70), RC (69.24 $\pm$ 11.74), NF (91.82 $\pm$ 20.58), $p = 0.007$ for RC vs NF and $p = 0.0003$ for RV vs NF; phase 4 - RV (17.59 $\pm$ 5.57), RC (17.06 $\pm$ 3.37), NF (16.94 $\pm$ 5.89)). Overall, the combination of continuous and vibrotactile (V) feedback yielded superior performance than (C), with participants achieving more accurate and consistent hand poses than with continuous feedback alone.

8.3 Conclusions

This chapter presented two wearable haptic devices for distributed feedback at the wrist and forearm, with the main design objective of being compact, lightweight, and comfortable for the user. At the same time, alternative feedback actuation methods were adopted with respect to the conventional vibrotactile stimuli generated by vibrating mass motors. The shallow torques approach aimed at a more directional and natural perception, closer to the sensation of an external guidance sensation (i.e. as by a therapist), still resulting in a low weight and compact implementation. The haptic armband instead explored the direct-drive approach, hence reducing the absolute intensity of the signal yet increasing the bandwidth (from continuous to fast dynamics) and the variety of rendered signals.

Experiments showed that the wristband with four servo modules provided clear and effective directional cues for wrist movements, significantly improving accuracy and

reducing variability, with performance comparable to visual feedback. The armband delivered continuous and vibrotactile stimuli that supported complex motor tasks like grasping and prono-supination. The combined feedback mode proved effective in guiding precise and consistent movements within a VR rehabilitation scenario.

Overall, these devices and experimental results demonstrate the potential of distributed wearable haptics to enhance motor control in immersive VR rehabilitation scenarios. However, the small sample size limits the generalization of the results, and future studies are aimed at involving larger groups of participants. Further experimental activities related to this technology are included in the validation of the Case Studies "Extended Reality for Rehabilitation" (Chapter 20) and "Extended Reality for People with Serious Mobility and Verbal Communication Diseases" (Chapter 22).

References

Aggravi, Marco, Florent Pausé, Paolo Robuffo Giordano, and Claudio Pacchierotti (2018). "Design and evaluation of a wearable haptic device for skin stretch, pressure, and vibrotactile stimuli". In: *IEEE Robotics and Automation Letters* 3.3, pp. 2166–2173.

Bark, Karlin, Jason Wheeler, Pete Shull, Joan Savall, and Mark Cutkosky (2010). "Rotational skin stretch feedback: A wearable haptic display for motion". In: *IEEE Transactions on Haptics* 3.3, pp. 166–176.

Bortone, Ilaria, Michele Barsotti, Daniele Leonardis, Alessandra Crecchi, Alessandra Tozzini, Luca Bonfiglio, and Antonio Frisoli (2020). "Immersive Virtual Environments and Wearable Haptic Devices in rehabilitation of children with neuromotor impairments: a single-blind randomized controlled crossover pilot study". In: *Journal of NeuroEngineering and Rehabilitation* 17.1, pp. 1–14.

Casini, Simona, Matteo Morvidoni, Matteo Bianchi, Manuel Catalano, Giorgio Grioli, and Antonio Bicchi (2015). "Design and realization of the cuff-clenching upper-limb force feedback wearable device for distributed mechano-tactile stimulation of normal and tangential skin forces". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1186–1193.

Chen, Yu, Kingsley Travis Abel, John T Janecek, Yunan Chen, Kai Zheng, and Steven C Cramer (2019). "Home-based technologies for stroke rehabilitation: A systematic review". In: *International journal of medical informatics* 123, pp. 11–22.

Chinello, Francesco, Claudio Pacchierotti, Joao Bimbo, Nikos G Tsagarakis, and Domenico Prattichizzo (2017). "Design and evaluation of a wearable skin stretch device for haptic guidance". In: *IEEE Robotics and Automation Letters* 3.1, pp. 524–531.

Dobbelstein, David, Evgeny Stemasov, Daniel Besserer, Irina Stenske, and Enrico Rukzio (2018). "Movelet: A self-actuated movable bracelet for positional haptic feedback on the user's forearm". In: *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pp. 33–39.

Eguchi, Ryo, David Vacek, Cole Godzinski, and Allison M Okamura (2023). "Between-tactor display using dynamic tactile stimuli for directional cueing in vibrating environments". In: *IEEE Transactions on Haptics*.

Frisoli, Antonio and Daniele Leonardis (2024). "Wearable haptics for virtual reality and beyond". In: *Nature Reviews Electrical Engineering* 1.10, pp. 666–679.

Koslover, Rebecca L, Brian T Gleeson, Joshua T De Bever, and William R Provancher (2011). "Mobile navigation using haptic, audio, and visual direction cues with a handheld test platform". In: *IEEE Transactions on Haptics* 5.1, pp. 33–38.

Levin, Mindy F, Patrice L Weiss, and Emily A Keshner (2015). "Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles". In: *Physical therapy* 95.3, pp. 415–425.

Meli, Leonardo, Irfan Hussain, Mirko Aurilio, Monica Malvezzi, Marcia K O'Malley, and Domenico Prattichizzo (2018). "The hBracelet: a wearable haptic device for the distributed mechanotactile stimulation of the upper limb". In: *IEEE Robotics and Automation Letters* 3.3, pp. 2198–2205.

Moriyama, Taha K, Ayaka Nishi, Rei Sakuragi, Takuto Nakamura, and Hiroyuki Kajimoto (2018). "Development of a wearable haptic device that presents haptics sensation of the finger pad to the forearm". In: *2018 IEEE Haptics Symposium (HAPTICS)*. IEEE, pp. 180–185.

Provancher, William (2014). "Creating greater VR immersion by emulating force feedback with ungrounded tactile feedback". In: *IQT Quarterly* 6.2, pp. 18–21.

Scheggi, Stefano, Agostino Talarico, and Domenico Prattichizzo (2014). "A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback". In: *22nd Mediterranean conference on control and automation*. IEEE, pp. 20–23.

Spiers, Adam J and Aaron M Dollar (2016). "Design and evaluation of shape-changing haptic interfaces for pedestrian navigation assistance". In: *IEEE transactions on haptics* 10.1, pp. 17–28.

Vitense, Holly S, Julie A Jacko, and V Kathlene Emery (2003). "Multimodal feedback: an assessment of performance and mental workload". In: *Ergonomics* 46.1-3, pp. 68–87.

9. XR Collaboration and Gaze-Based Interaction

Froso Sarri[1], *George Ramiotis*[1], *and Katerina Mania*[1]

[1] School of Electrical & Computer Engineering, Technical University of Crete (TUC), Greece

Abstract. Gaze-based systems are being explored to enhance user experiences in Extended Reality (XR) across applications, from collaboration to direct interaction. One area of research focuses on AR-enabled collaboration, where gaze cues can provide clear, non-verbal indicators of intent to reduce ambiguity. Our study investigated the use of gaze cues within guidance contexts, developing a system with dynamic gaze visualizations to deliver task-related guidance. When compared to traditional verbal instructions, this gaze-based system was found to significantly reduce errors and unnecessary visual exploration. It also improved collaborators' attention to task-relevant areas and lowered their perceived workload, proving effective for accelerating task completion. A separate field of investigation addresses challenges in gaze-based selection for hands-free interaction in XR environments. Methods like gaze dwell suffer from a trade-off between the Midas Touch problem (from small thresholds) and eye fatigue (from large thresholds). To solve this, we introduce CONTEXT-GAD, a novel context-aware adaptive dwell system that infers cognitive load by leveraging the task's visual context and user behavioral features. A hierarchical machine learning model then dynamically adapts dwell thresholds in three levels based on this load. A user study demonstrated that this adaptive approach significantly reduced task completion time in less complex tasks and improved error rates in more cognitively intensive scenes, enhancing efficiency and accuracy without increasing the perceived workload.

9.1 Introduction

Within Extended Reality (XR), the gaze modality is a critical subject of research, enabling new applications, including collaborative work and hands-free interactions.

One area of investigation, presented in this first study, focuses on the effect of gaze-based cues in co-located AR collaboration where one user guides another. While prior work has shown the benefits of shared gaze cues for improving mutual awareness and task understanding in remote collaboration [Bai et al. 2020; Jing et al. 2022a], researchers are now examining co-located setups to address challenges in aligning attention and intention without relying on verbal or physical cues [Jing et al. 2021]. Recent systems have begun visualizing gaze behavior in co-located setups, leveraging HMD's eye-tracking capabilities, enabling shared gaze cues via dynamic indicators like gaze rays, cursors, or changes in color and shape to signal attention and intent [D'angelo and Schneider 2021; Jing et al. 2022b]. These visualizations have improved joint attention and task coordination, with bi-directional cues and gaze-triggered designs proving particularly effective [Jing et al. 2021]. However, most existing studies assume knowledge symmetry between collaborators. Furthermore, most system evaluations rely on task-level metrics such as completion time and error rate [Brägger et al. 2022; Chen et al. 2021], or subjective measures, without deeper analysis of gaze behavior or attention patterns. Deeper gaze behavior analysis has either focused on non-AR settings [Acar et al. 2024] or used gaze passively rather than as a primary communication cue [Pathmanathan et al. 2024]. To address this gap, we developed a co-located AR system that uses dynamic gaze visualizations, rendered as a ray and pointer, to guide object placement tasks. We compared this approach to traditional verbal instructions using task performance metrics, gaze data, and subjective workload ratings. The goal was to evaluate whether gaze-based cues offer a more effective means of communicating task-related information in asymmetrical collaboration, specifically in terms of task efficiency and user behavior. By analyzing both visual attention patterns and task outcomes, we provide insights into the role of gaze as a primary communication modality in co-located AR guidance.

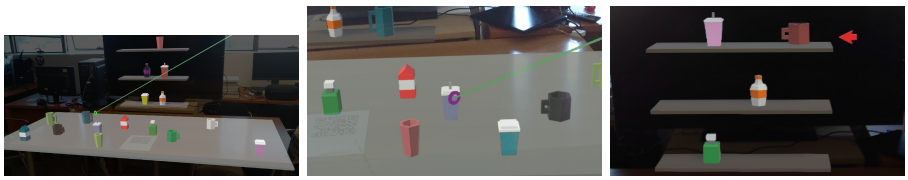
In a separate but related domain, the increasing adoption of XR headsets has driven research into hands-free interaction methods, with gaze-based selection being one of the most common [Argelaguet and Andujar 2013]. Dwell selection, where a user fixates on a target for a fixed duration, is a popular technique [Penkar et al. 2012]. However, its static nature presents a fundamental challenge: a short, fixed threshold is efficient but prone to unintentional activations, known as the Midas Touch problem, while a long threshold reduces errors but increases eye fatigue and interaction time. Prior research has attempted to solve this by adapting the dwell time. Some approaches

use probabilistic models based on past interactions [Panwar et al. 2012; Pi et al. 2020] or machine learning models to predict user intent based on gaze-only features [Narkar et al. 2024]. However, these methods are often heavily biased towards an individual user's unique gaze patterns or interaction strategies. A significant limitation is their failure to incorporate the visual context of the scene—such as the density, size, and location of surrounding objects—which can significantly influence a user's cognitive load and gaze behavior. In this chapter, we propose CONTEXT-GAD, a novel context-aware machine learning model for adapting dwell thresholds in XR. Our system dynamically adjusts the fixation time in three levels by leveraging both visual context features from the environment and behavioral features from the user to determine the perceived cognitive load of the scene. This approach allows the system to generalize across different users and scenarios, providing a more robust solution to the limitations of static dwell.

9.2 XR collaboration

9.2.1 Methodology and Results

In the developed system, two users wearing Microsoft HoloLens 2 devices collaborated in a shared environment where one acted as the instructor and the other performed object placement. The instructor had to select 8 out of 16 available 3D objects and guided placement onto virtual shelves using their gaze, visualized as a ray that ends to a pointer (Figure 9.1a) that changed size and color after 0.8s of fixation (Figure 9.1b). A 3D arrow annotation could also be placed using a gaze-pinch gesture (Figure 9.1c). The user had to pick the 3D object that the instructor had selected with the pointer and place it in the annotated shelf. We compared three conditions: Constant Gaze (CG), where the instructor's gaze visualization was always visible; Triggered Gaze (TG), where it appeared only during object fixation; and Verbal Instructions (VI), which used



(a) (b) (c)

Figure 9.1: (a) Instructor's gaze visualized in the task space. (b) Visualization of the instructor's gaze fixation. (c) 3D arrow placed on a shelf by the instructor to annotate it.

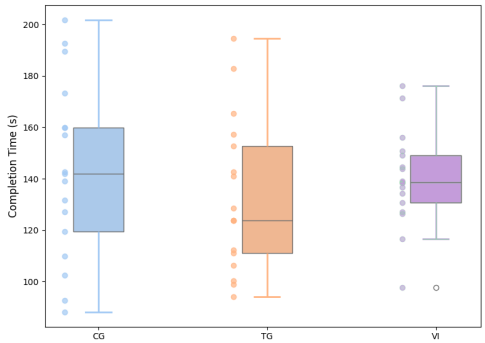


Figure 9.2: Box plots of task completion times of the three conditions, with individual data points.

verbal guidance only. Eye-tracking data were utilized to measure fixation durations on task-relevant Areas of Interest (AOIs): the 3D objects in the horizontal plane, the shelves in the vertical plane, and the collaborator.

User Study

The study involved 17 participants (10 female, 7 male, ages 18–44) with varying AR experience, using a within-subject design. Quantitative metrics include: task performance, which was measured by completion time and object interaction correctness, and gaze data, which comprises the relative fixation durations, which are calculated as the proportion of fixation time on each AOI relative to the overall task time. Subjective feedback includes NASA TLX and a general questionnaire to gather participant feedback on overall preference and their ratings on aspects of the conditions. A single instructor guided all sessions for consistency, and only the user's data was analysed because the user's performance, attention, and experience were the primary focus of the study.

Task Efficiency Results: The completion times of the three conditions are shown in Figure 9.2. The TG condition recorded the lowest mean completion time (μ=132.9 s, σ=29.7 s), followed by VI (μ=139.8 s, σ=18.7 s). The CG condition had the highest mean completion time (μ=142.8 s, σ=34.4 s). Repeated-measures ANOVA revealed no statistically significant differences between the conditions ($F_{2,32} = 0.734, p = 0.488, \eta_g^2 = 0.022$). Correctness scores can be seen on Table 9.1. A repeated-measures ANOVA showed a statistically significant effect of condition on correctness ($F_{2,32} = 12.82, p < 0.001, \eta_g^2 = 0.299$). Post-hoc pairwise comparisons (Bonferroni-corrected) showed that

Table 9.1: Mean and Standard Deviation of Object Interaction Correctness.

Condition	Mean (%)	SD (%)
CG	97.05	9.41
TG	97.05	5.47
VI	80.88	17.74

correctness scores for the CG and TG conditions were significantly higher than those for the VI condition ($p < 0.001, p = 0.011$, respectively). However, no significant difference was observed between CG and TG ($p = 1.000$). Considering the relationship between task completion time and correctness, while the CG condition recorded a higher mean completion time compared to the VI condition, it also achieved higher correctness scores than the VI condition. This suggests that the ambiguity of verbal instructions in the VI condition has led participants to make more errors despite completing the task more quickly. The TG condition demonstrated the best overall balance, achieving the lowest mean completion time while maintaining a high correctness score.

Gaze-Based Metrics: The results are presented in Figure 9.3. In the VI condition, the instructor was not immersed in the AR environment; therefore, no fixation durations were recorded for that AOI in that condition. Across the three conditions, participants spent the highest gaze time on the 3D objects, CG ((μ=29.74 %, σ=8.22 %)), TG (μ=30.5 %, σ=8.87 %), VI (μ=42.06 %, σ=9.65 %). Relative fixations on shelves were overall lower, CG ((μ=16.92 %, σ=6.03 %)), TG (μ=13.23 %, σ=3.74 %), VI (μ=14.03 %, σ=4.55 %). As was observed during the experiments, participants rarely looked at the instructor during the task, CG ((μ=0.58 %, σ=0.71 %)), TG (μ=1.03 %, σ=0.78 %).

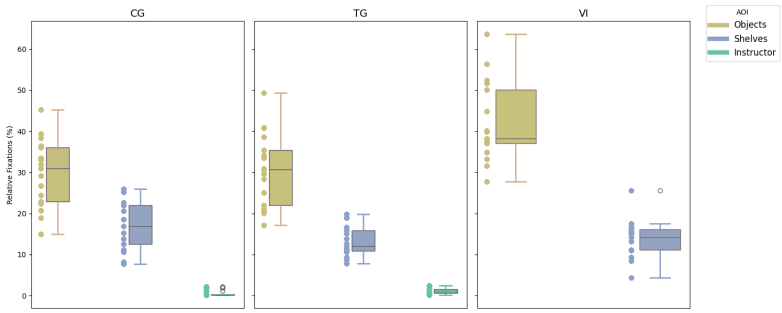


Figure 9.3: Box plots of relative fixation durations (%) on each defined AOI across the three conditions, with individual data points.

We excluded the instructor AOI from the CG and TG conditions in order to have a balanced repeated-measures ANOVA design, and it revealed significant main effects of both condition ($F_{5,80} = 54.72, p < 0.001, \eta_p^2 = 0.77$) and AOI ($F_{1,16} = 89.59, p < 0.001, \eta_p^2 = 0.85$), as well as significant interaction between condition and AOI ($F_{5,80} = 58.07, p < 0.001, \eta_p^2 = 0.78$). Post-hoc pairwise comparisons (Bonferroni-corrected) showed that fixations on the Objects were significantly higher in the VI condition than both the CG ($p < 0.001$) and TG ($p < 0.001$) conditions. Similarly, for the Shelves AOI, fixation durations were significantly lower in the TG condition compared to both CG ($p < 0.001$) and VI ($p < 0.001$). These results indicate that the VI condition resulted in the highest visual exploration times of the horizontal task space, as participants spent more time searching for the correct object without the immediate guidance provided by visual indications, such as those in the CG and TG conditions. For the Shelves AOI, fixation durations were lower in the TG condition compared to CG and VI.

User Feedback: Average NASA TLX scores across the three conditions are presented in Figure 9.4. Gaze-based conditions generally led to lower mental demand and frustration compared to VI, suggesting reduced cognitive load and ambiguity. CG had the lowest mental demand (μ=2.94, σ=1.25), followed by TG (μ=3.06, σ=1.60), while VI had the highest (μ=3.29, σ=1.86). Frustration was also lowest in CG (μ=2.94, σ=2.24), slightly higher in TG (μ=3.12, σ=2.44), and highest in VI (μ=3.76, σ=2.08). Interestingly, VI showed the lowest temporal demand (μ=3.65, σ=2.26) compared to TG (μ=3.88, σ=2.09) and CG (μ=4.11, σ=2.26), possibly due to the absence of time pressure from disappearing annotations in the gaze conditions. Physical demand was low across all conditions, with CG (μ=2.53, σ=1.28) and TG (μ=2.53, σ=1.77) slightly lower than VI (μ=2.88, σ=1.36). Effort ratings were comparable, with VI (μ=3.47, σ=2.00) slightly higher than CG (μ=3.29, σ=1.31) and TG (μ=3.24, σ=1.60). Performance was rated highest in CG and TG (μ=7.53, σ=3.02 and σ=2.76, respectively), compared to VI (μ=6.23, σ=2.68), indicating greater confidence and perceived effectiveness with gaze-based cues. Overall, the results suggest that gaze-based guidance reduced workload and improved task experience.

Regarding the user experience, participants generally favored the gaze-based conditions over verbal instructions. When asked if instructions were accurately represented to the users, CG received the most consistent high ratings across clarity and effectiveness of cues, with the majority strongly agreeing that instructions were accurately represented. TG also scored well but showed slightly more variability, likely due to the intermittent nature of its visualization. In contrast, VI received more mixed responses, with some participants reporting difficulty interpreting verbal-only guidance. When asked about overall preference, most participants selected CG (8), followed by TG (6),

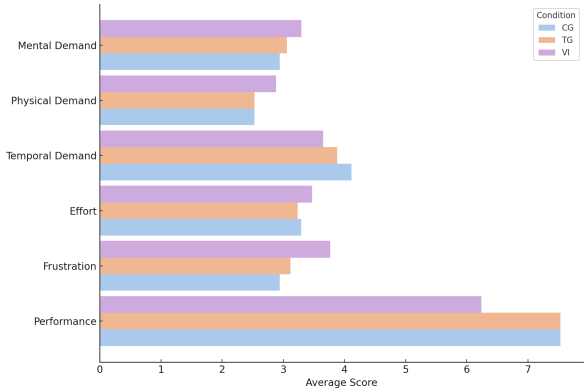


Figure 9.4: NASA TLX Scores by Condition (1-very low, 10-very high).

while only 3 preferred VI, suggesting that gaze-based cues were generally perceived as clearer, more engaging, and more supportive for task collaboration.

*This study was published in [Sarri and Mania 2025].

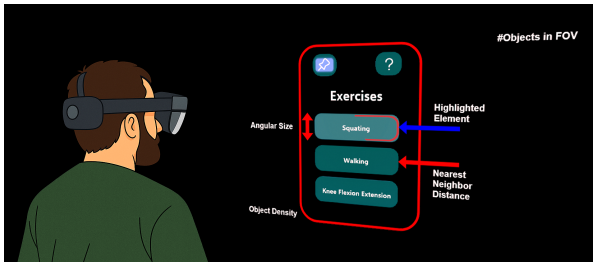
9.3 Gaze-based Interactions

9.3.1 Methodology and Results

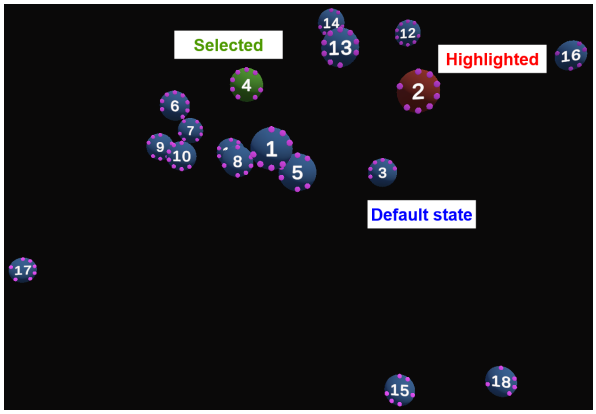
System and Data Collection

The system was developed in the Unity Engine using the Mixed Reality Toolkit (MRTK3) and deployed on a Microsoft HoloLens 2. To train our model, we first conducted a data collection experiment with 20 participants. Participants performed gaze-based interactions in two distinct scenarios: a general User Interface (UI) navigation task (Figure 9.5a) and a visual search task (Figure 9.5b).

A pilot study performed on the visual search and navigation task established a default dwell threshold of 1.5s. During the main experiment, after each interaction, participants were prompted to provide ground-truth feedback by selecting one of three options: 'Same' (if the 1.5s default felt appropriate), 'Shorter' (if the interaction could be faster), or 'Longer' (if they needed more time or a false activation occurred).



(a) The elements shown are: (1) the currently highlighted target, designated by a blue arrow; (2) a graphical representation of a subset of the contextual features (red box and arrows) that are analyzed in real time to estimate cognitive load; and (3) a progress indicator (red perimeter bar) that provides feedback to the user on the dwell activation

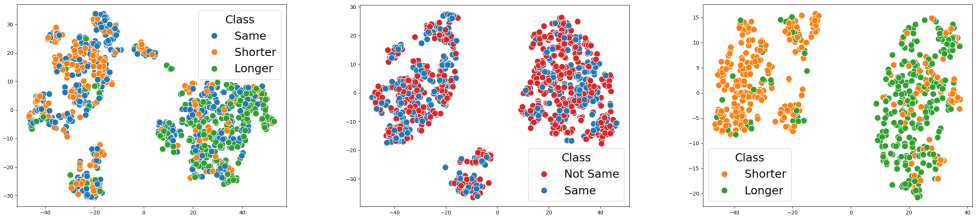


(b) The image highlights three distinct interaction states: 'Selected' (green sphere), 'Highlighted' (red sphere), and 'Default state' (blue spheres). Each sphere features colored dots along its perimeter.

Figure 9.5: An illustration of the general UI navigation task (a) and the visual search task (b). Figures adapted from [Ramiotis and Mania 2025] licensed under CC-BY 4.0.

Table 9.2: Description of collected features

Feature Name	Description	Units
Objects	The number of objects in the scene	Numerical
ObjectsFOV	The number of objects in the Field of View (FOV) of the user	Numerical
ObjectDensity	The density of the objects in the scene	$\text{objects}\,\text{m}^{-2}$
DynamicObjects	The number of dynamic objects	Numerical
NearestNeighborDistance	The distance of the nearest neighbor to the target	m
InteractFreq	The frequency of gaze interactions	$\text{interactions}\,\text{min}^{-1}$
GazeSpeedVariance	The variance of gaze speeds	$^{\circ 2}\,\text{s}^{-2}$
HeadRotVel	The rotational velocity of the head	$^{\circ}\,\text{s}^{-1}$
AngularSize	The angular size of the target	$^{\circ}$
LabelComp	The complexity of the target label	CEFR score



(a) Original multiclass problem

(b) Hierarchical first level (Same vs Not Same)

(c) Hierarchical second level (Shorter vs Longer)

Figure 9.6: t-SNE Projection of Features. Figures adapted from [Ramiotis and Mania 2025] licensed under CC-BY 4.0.

Feature Extraction and Classification

Our model's key innovation is its feature set (Table 9.2), which captures the cognitive demands of the scene. We extracted 10 features based on the reviewed literature on cognitive load. These features are divided into:

- *Visual Context Features:* Such as the number of objects in the scene [Doyon-Poulin et al. 2012] and in the user's Field of View (FOV) [Reis et al. 2012], 'ObjectDensity' [Whitney and Levi 2011], 'NearestNeighborDistance' (to account for visual crowding) [Hebbar et al. 2023], and the 'AngularSize' of the target.
- *Behavioral Features:* Such as 'InteractFreq' (frequency of interactions) [Vulpe-Grigorasi et al. 2024], 'GazeSpeedVariance' [Pillai et al. 2022], and 'HeadRotVel' (head rotation velocity), which act as indicators of user behavior under cognitive load.

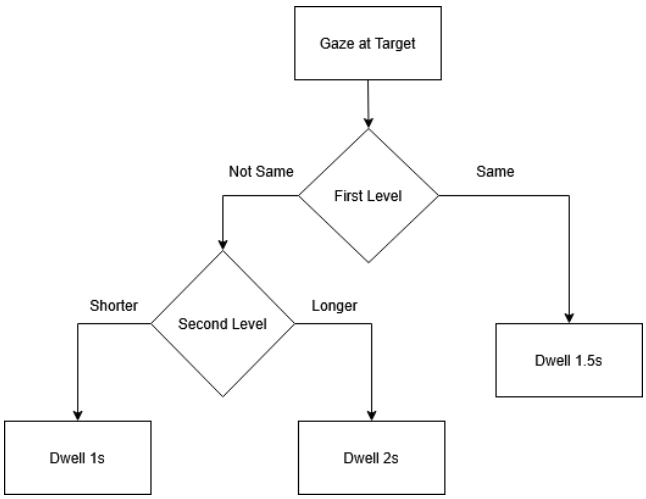


Figure 9.7: Hierarchical classifier structure for adaptive dwell times. Figure adapted from [Ramiotis and Mania 2025] licensed under CC-BY 4.0.

A two-level hierarchical classifier (Figure Figure 9.6) using the Random Forest algorithm was trained on the collected data (1385 samples). This structure, illustrated in Figure 9.7, was chosen because it outperformed a standard multiclass model.

- *Level 1:* A binary classifier distinguishes between 'Same' and 'Not Same' (i.e., 'Shorter' or 'Longer'). This level achieved an accuracy of **70.72%**.
- *Level 2:* If the first level predicts 'Not Same', a second binary classifier activates to distinguish between 'Shorter' and 'Longer'. This level achieved an accuracy of **85.43%**.

Evaluation Procedure

We conducted a separate user study with 17 participants to evaluate the trained system. A within-subjects design was used, where participants completed both the navigation and visual search tasks using two different systems:

1. *Static System:* Used a fixed 1.5s dwell threshold for all interactions.
2. *Adaptive System:* Used our trained hierarchical model to dynamically adjust the dwell time to be shorter (1s), same (1.5s), or longer (2s).

To prevent bias, the system order was counterbalanced, and participants were unaware of which system they were using. We measured objective metrics (task completion

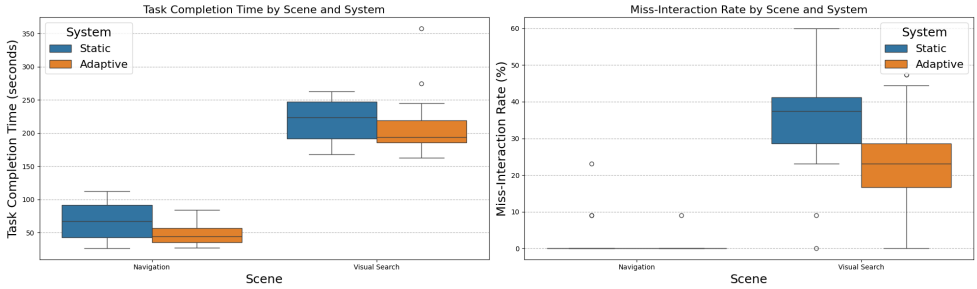


Figure 9.8: Quantitative results for each scene with each system, showing mean and standard error for (Left) Task Completion Time and (Right) Miss-interaction Rate. Figure adapted from [Ramiotis and Mania 2025] licensed under CC-BY 4.0.

time and miss-interaction rate) and subjective metrics (System Usability Scale (SUS) and NASA-TLX for workload).

Results and Discussion

The quantitative results (Figure 9.8) were analyzed using a non-parametric Friedman test, since the data were not normally distributed, followed by post-hoc Wilcoxon signed-rank tests with Bonferroni correction. The analysis revealed a significant and nuanced advantage for the Adaptive system.

- *Task Completion Time:* In the simpler navigation task, the Adaptive system was significantly faster than the Static system ($W = 17.00, p_{corrected} = 0.006$). In the complex visual search task, there was no significant difference in time.
- *Miss-Interaction Rate:* In the cognitively demanding visual search task, the Adaptive system resulted in a significantly lower miss-interaction rate ($W = 10.00, p_{corrected} = 0.009$). This demonstrates its effectiveness in mitigating the Midas Touch problem under high cognitive load. No difference was found in the simpler navigation task.

Subjective feedback from the SUS and NASA-TLX questionnaires reinforced these findings.

- *Usability (SUS):* The Adaptive system was rated as significantly more usable than the Static system, but only in the more demanding visual search task ($p_{corrected} = 0.02$).

- *Workload (NASA-TLX):* Critically, there was no significant difference in perceived workload between the two systems in either task. This indicates that the usability improvements and error reductions from the adaptive system were achieved without imposing any additional cognitive load on the user.

*This study was published in [Ramiotis and Mania 2025].

9.3.2 Conclusions

A co-located AR collaboration system was developed using gaze as the primary communication cue for task-related guidance. The study demonstrated the advantages of gaze-based cues over verbal commands, with both CG and TG conditions improving task correctness, reducing ambiguity, and lowering cognitive workload. While CG offered consistent visual guidance and was preferred overall, its constant visibility occasionally caused distractions. TG, in contrast, provided timely cues that balanced effectiveness with reduced visual clutter. Gaze data analysis showed that CG and TG effectively directed participants' attention to task-relevant areas, whereas in VI, participants exhibited longer visual exploration times. VI results highlighted the limitations of relying solely on verbal instructions for disambiguation.

In parallel, a separate study introduced a novel context-aware adaptive gaze dwell system that effectively addresses the limitations of static dwell thresholds in XR. By dynamically adjusting dwell times based on cognitive load—inferred from the scene's visual context and user behavior—this system provides a more robust and generalizable solution than methods based on user-specific gaze patterns. The evaluation demonstrated that the adaptive system improves task efficiency (speed) in simpler scenarios and, more importantly, improves interaction accuracy (fewer errors) in complex, cognitively demanding scenarios. It successfully mitigates the Midas Touch problem when it matters most, enhancing usability without increasing the user's perceived workload. This context-aware approach represents a significant step toward more intelligent, reliable, and user-friendly hands-free interactions in XR environments.

References

Acar, Ayberk, Jumanh Atoum, Amy Reed, Yizhou Li, Nicholas Kavoussi, and Jie Ying Wu (2024). "Intraoperative gaze guidance with mixed reality". In: *Healthcare Technology Letters* 11.2-3, pp. 85–92.

Argelaguet, Ferran and Carlos Andujar (2013). "A survey of 3D object selection techniques for virtual environments". In: *Computers & Graphics* 37.3, pp. 121–136.

Bai, Huidong, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst (2020). "A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing". In: *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13.

Brägger, Luca, Louis Baumgartner, Kathrin Koebel, Joe Scheidegger, and Arzu Çöltekin (2022). "Interaction and visualization design considerations for gaze-guided communication in collaborative extended reality". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4, pp. 205–212.

Chen, Lei, Yilin Liu, Yue Li, Lingyun Yu, BoYu Gao, Maurizio Caon, Yong Yue, and Hai-Ning Liang (2021). "Effect of Visual Cues on Pointing Tasks in Co-Located Augmented Reality Collaboration". In: *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*. SUI '21. Virtual Event, USA: Association for Computing Machinery.

D'angelo, Sarah and Bertrand Schneider (2021). "Shared gaze visualizations in collaborative interactions: Past, present and future". In: *Interacting with Computers* 33.2, pp. 115–133.

Doyon-Poulin, Philippe, Benoit Ouellette, and Jean-Marc Robert (Oct. 2012). "Review of visual clutter and its effects on pilot performance: New look at past research". In: *2012 IEEE/AIAA 31st Digital Avionics Systems Conference (DASC)*, pp. 1–36.

Hebbar, Archana, Sanjana Vinod, A. K. Shah, Abhay Pashilkar, and Pradipta Biswas (July 2023). "Cognitive load estimation in VR flight simulator". In: *Journal of Eye Movement Research* 15.33.

Jing, Allison, Kunal Gupta, Jeremy McDade, Gun A Lee, and Mark Billinghurst (2022a). "Comparing gaze-supported modalities with empathic mixed reality interfaces in remote collaboration". In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp. 837–846.

Jing, Allison, Kieran May, Gun Lee, and Mark Billinghurst (2021). "Eye See What You See: Exploring How Bi-Directional Augmented Reality Gaze Visualisation Influences Co-Located Symmetric Collaboration". In: *Frontiers in Virtual Reality* 2, p. 79.

Jing, Allison, Kieran May, Brandon Matthews, Gun Lee, and Mark Billinghurst (2022b). "The Impact of Sharing Gaze Behaviours in Collaborative Mixed Reality". In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2, pp. 1–27.

Narkar, Anish S., Jan J. Michalak, Candace E. Peacock, and Brendan David-John (May 2024). "GazeIntent: Adapting Dwell-time Selection in VR Interaction with Real-time Intent Modeling". In: *Proc. ACM Hum.-Comput. Interact.* 8.ETRA, 226:1–226:18.

Panwar, Prateek, Sayan Sarcar, and Debasis Samanta (2012). "EyeBoard: A fast and accurate eye gaze-based text entry system". In: *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*. IEEE, pp. 1–8.

Pathmanathan, Nelusa, Tobias Rau, Xiliu Yang, Aimée Sousa Calepso, Felix Amtsberg, Achim Menges, Michael Sedlmair, and Kuno Kurzhals (2024). "Eyes on the Task: Gaze

Analysis of Situated Visualization for Collaborative Tasks". In: *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 785–795.

Penkar, Abdul Moiz, Christof Lutteroth, and Gerald Weber (2012). "Designing for the eye: design parameters for dwell in gaze interaction". In: *Proceedings of the 24th Australian Computer-Human Interaction Conference*. OzCHI '12. Melbourne, Australia: Association for Computing Machinery, pp. 479–488.

Pi, Jimin, Paul A. Koljonen, Yong Hu, and Bertram E. Shi (Oct. 2020). "Dynamic Bayesian Adjustment of Dwell Time for Faster Eye Typing". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.10, pp. 2315–2324.

Pillai, Prarthana, Balakumar Balasingam, Yong Hoon Kim, Chris Lee, and Francesco Biondi (Aug. 2022). "Eye-Gaze Metrics for Cognitive Load Detection on a Driving Simulator". In: *IEEE/ASME Transactions on Mechatronics* 27.4, pp. 2134–2141.

Ramiotis, George and Katerina Mania (2025). "CONTEXT-GAD: A Context-Aware Gaze Adaptive Dwell model for Gaze-based Selections in XR Environments". In: *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada*.

Reis, Helena M, Simone S Borges, Vinicius HS Durelli, Luis Fernando de S Moro, Anarosa AF Brandao, Ellen F Barbosa, Leônidas O Brandao, Seiji Isotani, Patricia A Jaques, and Ig I Bittencourt (2012). "Towards reducing cognitive load and enhancing usability through a reduced graphical user interface for a dynamic geometry system: An experimental study". In: *2012 IEEE International Symposium on Multimedia*. IEEE, pp. 445–450.

Sarri, Froso and Katerina Mania (2025). "Effect of Gaze Visualization on Task Efficiency and User Behavior for Guidance Scenarios in Co-Located AR Collaboration". In: *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 561–568.

Vulpe-Grigorasi, Adrian, Benedikt Gollan, and Vanessa Leung (2024). "Assessing Cognitive Load in Distraction and Task Switching: Implications for Developing Realistic Clinical XR Training". In: *Computer Graphics International Conference*. Springer, pp. 84–98.

Whitney, David and Dennis M. Levi (Apr. 2011). "Visual crowding: a fundamental limit on conscious perception and object recognition". en. In: *Trends in Cognitive Sciences* 15.4, pp. 160–168.

10. Task Optimization and Prioritization

Blanca Guerrero[1], *Jordi Almendros*[1], *Josefa Mula*[2], *Elena Pérez-Bernabeu*[2], *and Marta Guerrero*[2]

[1] Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València (UPV), Spain

[2] Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València, Alarcón 1, 03801, Alcoy, Alicante, Spain

Abstract. In a dynamic work environment, prioritising human tasks is crucial for increasing efficiency and performance, while optimising operational costs. This chapter discusses a task optimisation component developed within the SUN project. It focuses on technologies that merge physical and virtual worlds to enhance social interactions and human collaboration. Based on a literature review and scenario definitions, we designed the task optimisation component, which includes the PRIORI-XR algorithm. This algorithm prioritises tasks in industrial settings by considering various criteria, such as equipment status, work shifts and resource availability. The component is integrated with image recognition and eXtended Reality (XR) technologies to create an immersive visualisation system. This system provides operators with contextual information and selection options through gestural interaction to thereby reduce cognitive load and facilitate real-time decision making, ulltimately improving individual worker performance. The preliminary results indicate that XR implementation in task management enhances execution accuracy and shortens decision-making times.

10.1 Introduction

In today's manufacturing industry context, efficient management and prioritisation of tasks have become essential for maintaining productivity and operational effectiveness in production systems, especially when resources are limited [Yang et al. 2021]. The shift towards digital production systems, boosted by new technologies, is creating opportunities to address these challenges in new ways.

Traditionally, organisations have depended on conventional techniques to arrange and allocate tasks, including to-do lists, planning applications or project management software. Although these methods can be beneficial, they often fall short and can be mentally taxing due to the growing complexity of production processes and the large amounts of data that operators must handle. In this context, eXtended reality (XR) technologies represent progress by allowing information to be added to the physical surroundings of real and virtual environments to assist workers in real time.

Augmented Reality (AR) enhances human-machine interactions by overlaying virtual information on real-world environments, which improves manufacturing processes, reduces rework and provides context-sensitive information for various tasks [Ong et al. 2008]. Virtual Reality (VR) creates realistic digital environments for user immersion and interaction, often used in industrial sectors for manufacturing planning, ergonomics, product design assessment and employee training [Aurich et al. 2008]. Mixed reality (MR) combines AR and VR to reduce uncertainties by merging real and virtual elements. It offers spatial flexibility, lower safety risks and improved robot interaction. MR systems integrate real objects with digital elements like holograms using devices, such as Microsoft HoloLens for interactive experiences [Ostanin et al. 2020].

XR encompasses technologies, including AR, MR and VR, that merge physical and virtual worlds to enhance experiences along the reality-virtuality continuum [Gong et al. 2021]. It can transform company operations by enabling immersive data visualisation and intuitive interaction. So it can significantly improve task prioritisation, reduce errors and optimise execution times.

The Social and Human Centred XR (SUN) project aims to enhance social and human interactions by merging physical and virtual worlds through AI-driven 3D acquisition technologies to address challenges in XR adoption and create accurate digital twins [Vairo et al. 2023]. One of the application scenarios addressed by the SUN project is the prioritisation of tasks in industrial XR environments. This article focuses on the task optimisation component, which is responsible for ordering workers' activities according to criteria of urgency, relevance and environmental conditions. This component integrates with other system modules to form a complete XR experience by providing

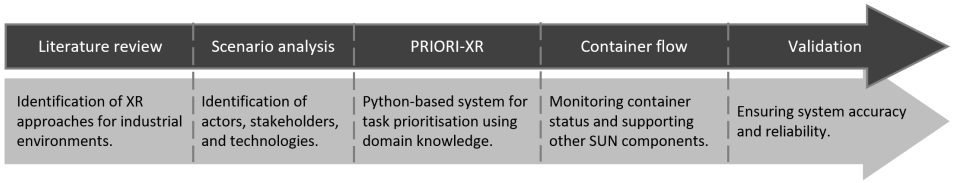


Figure 10.1: Task optimisation component methodology

contextual information and visual instructions to, thus, facilitateg operators' real-time decision making.

The ongoing structure of this chapter is organised as follows. Section 10.2 outlines the followed methodology. Section 10.3 provides a review of the background and related work about XR application in industrial contexts. Section 10.4 describes the scenario employed to develop the prioritisation component discussed in Section 10.5. Finally, Section 10.6 presents the conclusions and further research.

10.2 Metholodology

The development of the task optimisation component is structured into several stages (Figure 10.1). Firstly, a literature review explores existing approaches in XR for industrial environments. Next various scenarios are analysed to identify the key actors, stakeholders and involved technologies. Based on these insights, a rule-based logic system is implemented in Python to prioritise human tasks by leveraging its interpretability and ability to integrate domain knowledge. A container flow algorithm is also created to monitor container status and to support other SUN components. The process concludes with an internal validation phase to ensure system accuracy and reliability.

10.3 Literature Review

This section explores the challenges and opportunities of industrial XR technologies by focusing on the development of a human task prioritisation model. One of the SUN XR projects highlights concerns about XR applications for social interactions at work. The existing literature primarily focuses on the human-robot interaction, and leaves the human-human interaction in industrial settings largely unexplored, which presents a challenge for the project. Innovations like blockchain, digital twins and the Internet

of Things (IoT) enhance XR applications in manufacturing, particularly for maintenance training and assembly tasks [Doolani et al. 2020]. XR applications can train operators in machinery use, enhance workplace safety and create immersive experiences. They also support decision-making processes like forecasting, predictive maintenance, production planning, assembly and inventory management.

In the AR applications field in industrial environments, several significant contributions have emerged over the years. Li et al. create a virtual robot work cell for teaching that features view tracking with a virtual camera, visual and audio rendering, and a user interface [Li et al. 2000]. Zhong et al. develop a prototype for a collaborative industrial training system using AR [Zhong et al. 2003] . Tatić and Tešić propose an AR system to reduce training deficiencies and to minimise task monotony [Tatić and Tešić 2016]. More recently, Park et al. propose a user-centric task assistance method that integrates deep learning (DL), object detection and instance segmentation with wearable AR technology [Park et al. 2020]. Other strategies using AR and DL, in this case for predictive maintenance, are proposed by Liu et al. [Liu et al. 2022] and Wang et al. [Wang et al. 2022]. Lastly, Liu et al. develop an AR approach employing deep reinforcement learning (DRL) and cloud-edge orchestration for robot teaching [Liu et al. 2023]. DRL helps in robot motion planning, AR glasses enhance human-robot interaction, and cloud-edge orchestration facilitates communication between the AR platform and edge nodes.

VR is widely applied for training across industries, and emphasises sequential operations, procedures and risk assessments [Ji et al. 2022]. Programmes like that of Joshi et al. enhance safety in the concrete industry through training in personal protective equipment (PPE) and load management [Joshi et al. 2020]. Zhou et al. explore the human-robot interaction utilising VR and DL for robot teleoperation in civil engineering [Zhou et al. 2020]. Chadalavada et al. investigate the use of eye-tracking glasses as safety equipment by analysing navigation-related gaze data [Chadalavada et al. 2020] . Pérez et al. integrate VR and robotics for training and simulation [Pérez et al. 2019]. This aligns with [Tichon 2007], who note that VR can be used to simulate hazardous work conditions to help trainees to practice cognitive skills like problem solving and decision making.

For MR applications, Hasanzadeh, Polys and De La Garza investigate risk-taking behaviour in construction using passive haptics [Hasanzadeh et al. 2020]. XR applications have developed frameworks like SEEROB to assess ergonomics in robotic workstations [Weistroffer et al. 2022].

XR applications in industrial environments are still emerging, with limited literature available. There are opportunities for integrating these technologies with decision-support systems using operations research techniques like mathematical programming,

simulation and metaheuristics/matheuristics, alongside artificial intelligence (AI) technologies, such as reinforcement learning (RL) [Esteso et al. 2022], DRL [Liu et al. 2024], machine learning (ML) [Usuga et al. 2020] and DL to enhance XR applications while adhering to defined rules.

10.4 Scenario Definition

Shop floors can be hazardous due to the presence of machinery, metal rods and heavy containers filled with raw materials and waste from machines. Even though there are clearly marked walking and stationary areas, daily interactions frequently result in misplaced items, which increases the risk of accidents. Workers must take responsibility for minimising these risks while maintaining production by managing the status of containers and cutting metal rods.

To address these issues, a task optimisation component has been introduced to reduce stress and confusion by providing structure to the currently arbitrary appearance of tasks to be done by humans. This solution recognises the position and status of containers (both raw materials and waste), which will help to generate a prioritised task list. For example, the list may include tasks, such as emptying a specific container X in the waste area, filling container X with raw material, among others. Each task on the list will be ordered according to its priority.

Interaction with the system can be conducted via other components through hand input and gesture recognition. The technologies required for this solution include object recognition and location tracking, interactive streaming, application development, room scanning, a virtual machine for data storage and management, and cyber threat detection.

10.5 Task Optimization Component

The task optimisation component is developed to improve task management in a workshop and to balance its container status scenario. This component receives the necessary communications to run the task prioritisation algorithm, PRIORI-XR. Afterwards it transmits its decisions to the gesture recognition component. These decisions are then validated through XR. Figure 10.2 illustrates how the task optimisation component interacts with its submodules and other SUN components. The main components of the functioning system are highlighted in blue, while the required addons for validation, integration and evaluation purposes are shown in white.

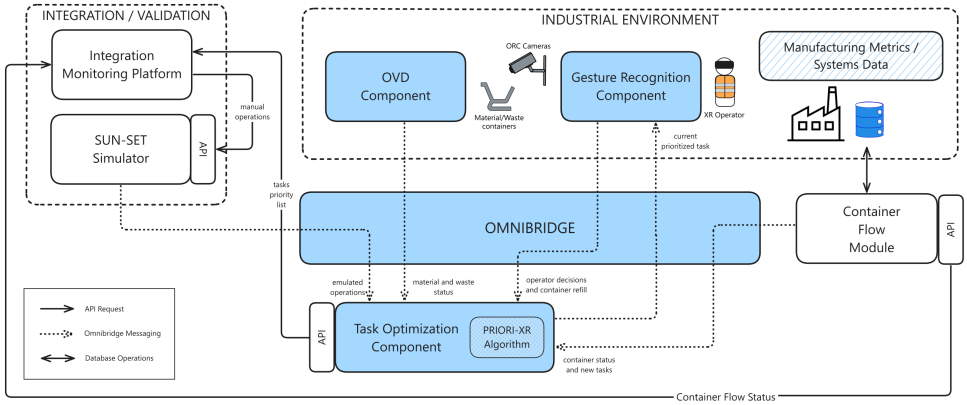


Figure 10.2: The task optimisation component architecture.

PRIORI-XR is a rule-based heuristic algorithm designed to manage the planning and allocation of tasks in industrial environments. Its operation is based on predefined criteria and the evaluation of contextual conditions to optimise factory operation through efficient resource allocation and orderly task execution.

The process starts with the loading of relevant data about the factory status, including information on machines, containers, raw materials, routes, work shifts and pending tasks. From these data, the PRIORI-XR algorithm determines the current shift (morning, afternoon or night) and filters out completed tasks. It then analyses the operational requirements and generates new tasks based on the industrial environment. Functions include the assignment of empty containers to create replenishment tasks, the balanced allocation of workloads between cutting machines according to their available capacity, the identification of full waste containers for emptying and the generation of cleaning tasks associated with shift changes. When the system receives a message from the open-vocabulary object detection (OVD) component – which processes RGB images and detects objects that match a textual description provided by the user, or the container flow module indicating either a raw material requirement or a filled waste container, see Chapter 5 – the task optimisation component determines whether a new task should be created. If necessary, the PRIORI-XR algorithm processes the task and generates a priority-ordered list. The task with the highest priority is then sent and displayed on the XR glasses for operator evaluation using the gesture recognition component to recognize custom hand gestures ("thumbs up" for "accept", "thumbs down" for "reject", and "moving two fingers" for "next") of the users based on the hand joints detected. Each task includes an identifier, description, priority and status, and tasks are organised to ensure efficient execution while avoiding redundancy or

duplication. When a decision about the highest priority task is communicated by the gesture recognition component, the task optimisation component makes a coherency evaluation to determine whether the decision remains valid in the current system state. If validated, the system updates the task status accordingly: tasks marked as accepted transition to progress; rejected tasks are deprioritised and moved down the task queue; completed tasks are removed from the active list. In the completion case, the system can also generate related tasks in accordance with predefined operational rules. In both the rejection and completion cases, PRIORI-XR is invoked again to re-evaluate task priorities.

All standard system communications take place through the Omnibridge communication solution, which acts as a central data broker, managing session-based communication and ensuring seamless interaction among XR applications, sensors, analytics engines, and other modules, using a standardised JSON format. Thus, OmniBridge is the middleware core of the SUN integrated XR platform (see Chapter 16), enabling secure, real-time communication and coordination between distributed XR components. When the task optimisation component receives a message through this channel, it activates internal data update procedures and starts the PRIORI-XR algorithm.

To facilitate the integration of the PRIORI-XR algorithm, the container flow module, the SUN-SET Simulator and the integration monitoring platform were created. All these submodules feature their own HTTP REST APIs and are designed as an optional, non-essential addition whose aim is to assist in validation and system integration.

The container flow module manages the materials flow in an industrial production environment by updating in real time the status of both raw material and swarf containers because parts are processed on different machines. It uses data from the Manufacturing Execution System (MES) to automatically discount the consumed raw material and to add the generated scrap as waste. When it detects that raw material falls below a predefined threshold or the waste chip is approaching the maximum container capacity, the system communicates this information to the prioritisation algorithm to generate operational tasks. This module acts as a support system for the OVD component, which is implemented by cameras in the plant to allow the visually captured information to be supplemented and/or validated with the MES data. This module is created as a support system for the OVD component because it may not be possible to install cameras on all the machines in a plant due to either physical limitations or data protection and industrial privacy issues.

The integration monitoring platform features a web interface that is accessible over the internal network and it showcases in real time all the system communications that pass through Omnibridge. It also allows for manual message injection through the SUN-SET Simulator, an additional component that has been developed to facilitate

the emulation of any component interaction for integration testing and non-real-time decision making. Moreover, it functions to monitor the prioritised tasks list and container status in real time. Hence, the SUN-SET Simulator has been developed to assess PRIORI-XR algorithm performance in diverse industrial scenarios. It functions autonomously and can transmit any type of message utilised in the system via Omnibridge to facilitate realistic interaction simulations in communication terms.

The system and their modules are composed of the following services:

Task Optimisation Component PRIORI-XR: the task prioritisation component that includes the PRIORI-XR algorithm. It prioritises the tasks associated with machine operations, is the main component of the communications, can work standalone, and has been Developed in Python;

Container Flow Module: manages raw material consumption and scrap generation on industrial machines. It Updates container status in real time and triggers information when certain thresholds are reached by generating tasks through the prioritisation algorithm. it has been developed in Python;

Integration APIs: flask-based REST APIs;

SUN-SET Simulator Module: emulates all the possible system actions through real communications as a component would do. It has been developed in Python for validation and integration, and enables manual operations on the system;

Integration Monitoring Platform: a Web interface for monitoring and controlling the system. Developed in Next.js framework and React, it connects through APIs to the simulator and component. Each module can be independently deployed. However, both the SUN-SET Simulator and the Integration Monitoring Platform require the Task Optimisation Component to be running first. Additionally, the Integration Monitoring Platform relies on the Simulator if manual system control is required.

Figure 10.3 illustrates how the Integration Monitoring Platform module allows for the in-depth control and the real-time monitoring of the system's current state. The updated code, along with the revised usage instructions, is available on `https://evia.in-two.com/sun/priori-xr-task-priorization-algorithm#`.

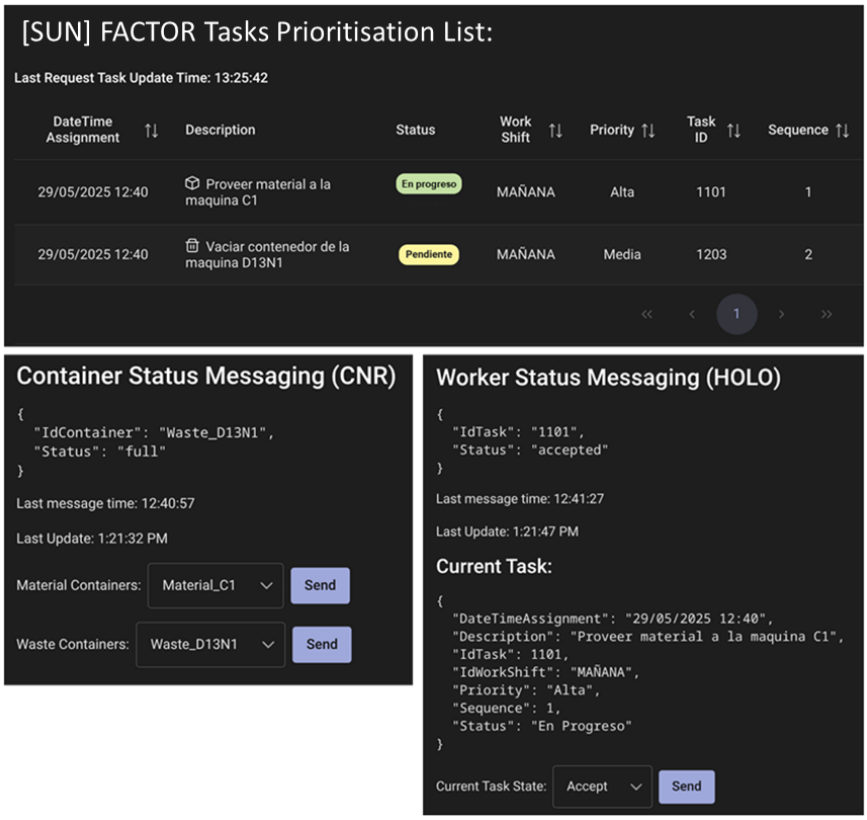


Figure 10.3: Integration Monitoring Platform. (Top) Current active task list. (Bottom left) Last communications with the OVD component. (Bottom right) Last received gesture message from the last send task from the gesture recognition component.

10.6 Conclusions

The design and implementation of a task optimisation component, integrated into the SUN project by combining rule-based logic with XR technologies to enhance task prioritisation in industrial environments, is presented. The developed PRIORI-XR algorithm prioritises human tasks by evaluating industrial environment data, such as equipment status, container levels, shift schedules and pending tasks. Integrated with a real-time visualisation system using image recognition and gesture-based XR interaction, the solution enables workers to receive, validate and execute tasks based on their urgency and operational impact. The preliminary validation shows that the component enhances decision making by reducing the cognitive load associated with manual task sorting, while

improving execution accuracy and shortening response times. The incorporation of XR glasses and gesture recognition facilitates intuitive interaction with task information by allowing human-system collaboration on the shop floor. This research contributes to address current gaps in task management systems and demonstrates the potential of XR to be implemented in industrial environments. Although XR implementation poses challenges, such as hardware constraints, integration costs and the need for robust data communication, its potential advantages regarding its usability are encouraging. Future work focuses on improving XR adaptability in dynamic manufacturing environments and further evaluating its impact across diverse industrial scenarios.

References

Aurich, J. C., D. Ostermayer, and C. H. Wagenknecht (2008). "Improvement of Manufacturing Processes with Virtual Reality-Based CIP Workshops". In: *International Journal of Production Research* 47.19, pp. 5297–5309.

Chadalavada, Ravi Teja, Henrik Andreasson, Maike Schindler, Rainer Palm, and Achim J. Lilienthal (Feb. 2020). "Bi-Directional Navigation Intent Communication Using Spatial Augmented Reality and Eye-Tracking Glasses for Improved Safety in Human–Robot Interaction". In: *Robotics and Computer-Integrated Manufacturing* 61 :101830.

Doolani, Sanika, Callen Wessels, Varun Kanal, Christos Sevastopoulos, Ashish Jaiswal, Harish Nambiappan, and Fillia Makedon (2020). "A Review of Extended Reality (XR) Technologies for Manufacturing Training". In: *Technologies* 8.4, p. 77.

Esteso, Ana, David Peidro, Josefa Mula, and Manuel Díaz-Madroñero (2022). "Reinforcement Learning Applied to Production Planning and Control". In: *International Journal of Production Research* 61.16, pp. 5772–5789.

Gong, Liang, Asa Fast-Berglund, and Bjorn Johansson (Jan. 2021). "A Framework for Extended Reality System Development in Manufacturing". In: *IEEE Access* 9, pp. 24796–24813.

Hasanzadeh, Sogand, Nicholas F. Polys, and Jesus M. De La Garza (2020). "Presence, Mixed Reality, and Risk-Taking Behavior: A Study in Safety Interventions". In: *IEEE Transactions on Visualization and Computer Graphics* 26.5, pp. 2115–2125.

Ji, Zuzhen, Yuchen Wang, Yinqiao Zhang, Yixuan Gao, Yi Cao, and Shuang-Hua Yang (Nov. 2022). "Integrating Diminished Quality of Life with Virtual Reality for Occupational Health and Safety Training". In: *Safety Science* 158 :105999.

Joshi, Sayali, Michael Hamilton, Robert Warren, Danny Faucett, Wenmeng Tian, Yu Wang, and Junfeng Ma (Oct. 2020). "Implementing Virtual Reality Technology for Safety Training in the Precast/ Prestressed Concrete Industry". In: *Applied Ergonomics* 90 :103286.

Li, Y. F., J. Ho, and N. Li (2000). "Development of a Physically Behaved Robot Work Cell in Virtual Reality for Task Teaching". In: *Robotics and Computer-Integrated Manufacturing* 16.2-3, pp. 91–101.

Liu, Changchun, Dunbing Tang, Haihua Zhu, Qingwei Nie, Wei Chen, and Zhen Zhao (Feb. 2024). "An Augmented Reality-Assisted Interaction Approach Using Deep Reinforcement Learning and Cloud-Edge Orchestration for User-Friendly Robot Teaching". In: *Robotics and Computer-Integrated Manufacturing* 85 :102638.

Liu, Changchun, Zequn Zhang, Dunbing Tang, Qingwei Nie, Linqi Zhang, and Jiaye Song (Mar. 2023). "A Mixed Perception-Based Human-Robot Collaborative Maintenance Approach Driven by Augmented Reality and Online Deep Reinforcement Learning". In: *Robotics and Computer-Integrated Manufacturing* 83 :102568.

Liu, Changchun, Haihua Zhu, Dunbing Tang, Qingwei Nie, Tong Zhou, Liping Wang, and Yejia Song (Apr. 2022). "Probing an Intelligent Predictive Maintenance Approach with Deep Learning and Augmented Reality for Machine Tools in IoT-Enabled Manufacturing". In: *Robotics and Computer-Integrated Manufacturing* 77 :102357.

Ong, S. K., M. L. Yuan, and A. Y. C. Nee (2008). "Augmented Reality Applications in Manufacturing: A Survey". In: *International Journal of Production Research* 46.10, pp. 2707–2742.

Ostanin, M., R. Yagfarov, D. Devitt, A. Akhmetzyanov, and A. Klimchik (2020). "Multi Robots Interactive Control Using Mixed Reality". In: *International Journal of Production Research* 59.23, pp. 7126–7138.

Park, Kyeong-Beom, Minseok Kim, Sung Ho Choi, and Jae Yeol Lee (June 2020). "Deep Learning-Based Smart Task Assistance in Wearable Augmented Reality". In: *Robotics and Computer-Integrated Manufacturing* 63 :101887.

Pérez, Luis, Eduardo Diez, Rubén Usamentiaga, and Daniel F. García (Aug. 2019). "Industrial Robot Control and Operator Training Using Virtual Reality Interfaces". In: *Computers in Industry* 109, pp. 114–120.

Tatić, Dušan and Bojan Tešić (Nov. 2016). "The Application of Augmented Reality Technologies for the Improvement of Occupational Safety in an Industrial Environment". In: *Computers in Industry* 85, pp. 1–10.

Tichon, Jennifer (2007). "Training Cognitive Skills in Virtual Reality: Measuring Performance". In: *CyberPsychology & Behavior* 10.2, pp. 286–289.

Usuga, Cadavid, Juan Pablo, Samir Lamouri, Bernard Grabot, Robert Pellerin, and Arnaud Fortin (2020). "Machine Learning Applied in Production Planning and Control: A State-of-the-Art in the Era of Industry 4.0". In: *Journal of Intelligent Manufacturing* 31.6, pp. 1531–1558.

Vairo, Claudio, Marco Callieri, Fabio Carrara, Paolo Cignoni, Marco Di Benedetto, Claudio Gennaro, Daniela Giorgi, Gianpaolo Palma, Lucia Vadicamo, and Giuseppe Amato (2023). *Social and human centered XR*. Ital-IA.

Wang, Liping, Dunbing Tang, Changchun Liu, Qingwei Nie, Zhen Wang, and Linqi Zhang (2022). "An Augmented Reality-Assisted Prognostics and Health Management System Based on Deep Learning for IoT-Enabled Manufacturing". In: *Sensors* 22.17, p. 6472.

Weistroffer, Vincent, François Keith, Arnaud Bisiaux, Claude Andriot, and Antoine Lasnier (2022). "Using Physics-Based Digital Twins and Extended Reality for the Safety and Ergonomics Evaluation of Cobotic Workstations. Frontiers in Virtual Reality 3 (February)".

Yang, Zhaojun, Jieli Li, Chuanhai Chen, Jialong He, Hailong Tian, and Lijuan Yu (Aug. 2021). "A Study on Overall Line Efficiency (OLE) Centered Production Line Maintenance Prioritization Considering Equipment Operational Reliability". In: *The International Journal of Advanced Manufacturing Technology* 124.11-12, pp. 3783–3794.

Zhong, X., P. Liu, N. D. Georganas, and P. Boulanger (2003). "Entwurf Einer Kollaborativen Augmented-Reality-Anwendung Für Industrielles Training (Designing a Vision-Based Collaborative Augmented Reality Application for Industrial Training)". In: *It - Information Technology* 45.1, pp. 7–19.

Zhou, Tianyu, Qi Zhu, and Jing Du (Sept. 2020). "Intuitive Robot Teleoperation for Civil Engineering Operations with Virtual Reality and Deep Learning Scene Reconstruction". In: *Advanced Engineering Informatics* 46 :101170.

Wearable and Vision-Based Monitoring Technologies for XR

The following part presents a suite of multimodal monitoring technologies that capture users' physical and emotional states through muscle, motion, gesture, and facial analysis. Integrating wearable sensors, computer vision, and machine learning, these systems collectively track posture, movement, intention, and emotions within the SUN XR platform. Together, they enable continuous observation of user interactions in Extended Reality, and provide real-time feedback, supporting users' awareness and more natural, adaptive, and engaging XR experiences.

11. EMG Decoding System for Hand and Wrist Kinematics

Aiden Xu[1], *Elena Ferrazzano*[2], *Vincent Mendez*[1,3], *and Silvestro Micera*[1,3,4],

[1] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[2] University of Cagliari, Italy,
[3] Centre Hospitalier Universitaire Vaudois, Switzerland
[4] Scuola Superiore Sant'Anna (SSSA), Italy

Abstract. The SUN XR platform enables people with limited mobility to navigate and socialize in virtual environments. User intention decoding plays an important role in the SUN XR platform, as it is the interface between the user and the virtual world. We developed a wearable, surface Electromyography(sEMG) based intention decoding system that decodes the user's navigation intentions from forearm muscle activities. The user's EMG signals are acquired by our custom 32-channel data acquisition system, streamed to a decoder network, and interpreted as navigation commands in the Virtual Reality (VR) world. The positive results (approximately 90% decoding accuracy) from our grasp classification tests in six subjects with Spinal Cord Injuries (SCI) show a promising use of this EMG-based intention decoding in a VR environment.

11.1 Introduction

The surface Electromyography (sEMG) is a common technique used to record and analyze muscle activities. Its non-invasive nature makes it favored in the fields of clinical assessment [Campanini et al. 2020; McManus et al. 2020], neurorehabilitation [Brucker and Buylaeva 1996; Nam et al. 2022; Pan et al. 2018], and motor decoding for prostheses [Kalbasi et al. 2024; Englehart and Hudgins 2003; Zhuang et al. 2019]. In

addition to the application in these fields, sEMG has also been applied to user-intention decoding in VR applications [Dwivedi et al. 2020].

The EMG signal acquisition from the forearm muscles is a component that allows the interpretation of subtle nuances of wrist and finger movements to translate these physiological signals into commands for interaction with Extended Reality (XR). A precise and easy-to-calibrate system is required to obtain a satisfactory user-device interface, particularly in contexts demanding high dexterity or for persons with motor impairment.

Together with the idea of sEMG-based intention decoding for VR applications, the sEMG decoding system described in this Chapter aims to improve the decoding capabilities, calibration times, and integration on a wearable device.

11.2 Methodology and Results

A medium-density (MD) EMG system prototype already existed at the start of SUN and was compared during the project against a gold-standard clinical system, optimizing deep networks to maximize decoding performance. The MD EMG system does not require specific placement of electrodes by a trained person compared to the gold standard one (Figure 11.1).

Clinical System Medium Density EMG

Figure 11.1: (Left) Picture of the gold standard EMG system used (Noraxon desk DTS). (Right) Picture of the Medium-Density EMG system. It consists of 64 monopolar electrodes placed around the forearm of the user.

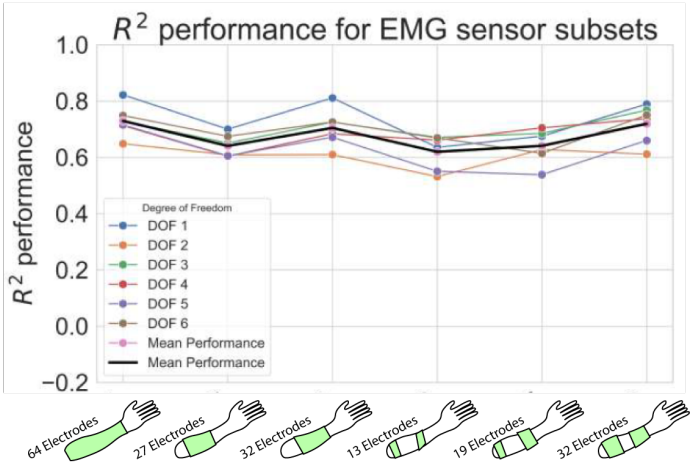


Figure 11.2: R2 single-finger decoding performance for each degree of freedom when considering EMG sensor subsets

Higher performance was obtained with the prototype, as well as a higher generalizability of the optimized deep model architectures. However, donning and doffing of the system needed to be simplified, and the EMG acquisition system had to be significantly miniaturized to improve wearability.

11.2.1 EMG Electrode Sleeve Improvements

In the process of improving the existing MD EMG system, we began by conducting an offline analysis of the existing 64-channel sleeve to evaluate the contribution of electrodes to the overall performance. Our analysis revealed that using only half of the electrodes provided similar performance to the full set of 64 electrodes. As a result, we decided to develop a new version of the system that utilizes just 32 electrodes. Various electrode subsets and placement options were considered during this redesign process (Figure 11.2). Ultimately, the simplest and most effective approach was to select the 32 most central electrodes (Figure 11.3).

The redesign also focused on reducing the setup time required for users to use the MD EMG system. With half as many electrodes as before and with an improved electrode layout, the setup process was significantly shortened, making the system much easier and quicker to deploy. Importantly, these improvements have been made without compromising the quality of the EMG signal decoding, ensuring that the streamlined system retains the same high level of performance as the original 64-channel version.

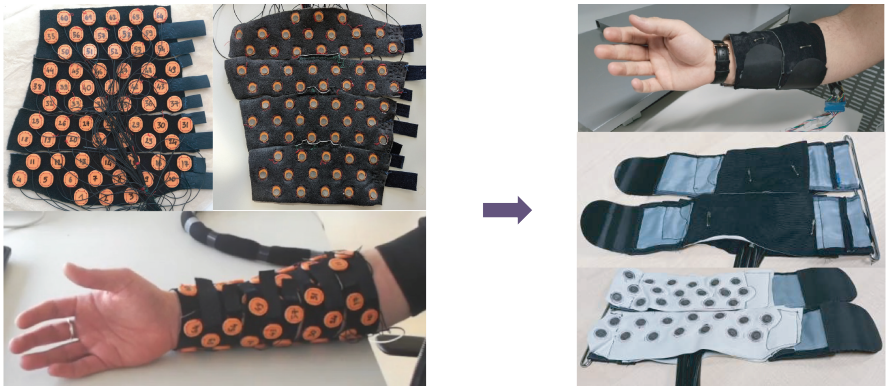


Figure 11.3: (Left) first prototype with 64 electrodes. (Right) second prototype with 32 electrodes.

11.2.2 EMG Acquisition Hardware Improvements

As the EMG electrode sleeve evolved, the EMG acquisition hardware also experienced several upgrades. From the clinical medium-density EMG system used in the early stage of the project, we designed three versions of portable EMG acquisition device to address the wearability issue of the clinical data acquisition system (Figure 11.4).

A wired prototype V0 that samples 32 channels was implemented as a first step toward miniaturization. After a performance benchmark, it was observed that prototype V0's signal-to-noise ratio was not enough to be used in a real-world scenario. We thus started prototype V1 to improve the signal-to-noise ratio and test wireless communication between the prototype and PC. Prototype V1 has adjustable gains from 1000X to more than 10000X. High common-mode rejection amplifiers and layout optimizations were employed to improve noise immunity and reduce interference. An onboard processor is added to simplify debug process and tests.

The latest and final version of the component, the EMG V3, is a 32-channel wireless EMG acquisition system. It features upgraded WiFi6 support for improved wireless transmission efficiency and reduced latency, while preserving Bluetooth Low Energy 5.3. When connected to devices with WiFi6 support, the EMG V3 can achieve robust transmission even in congested radio frequency environments. The previously proven successful time-multiplexed data acquisition architecture has been held in V3. A dedicated electrophysiology amplifier is now integrated into the V3 for a higher signal-to-noise ratio and up to 32-channel acquisition. The V3 has been integrated with our latest second-generation 32-channel electrode sleeve that was optimized for

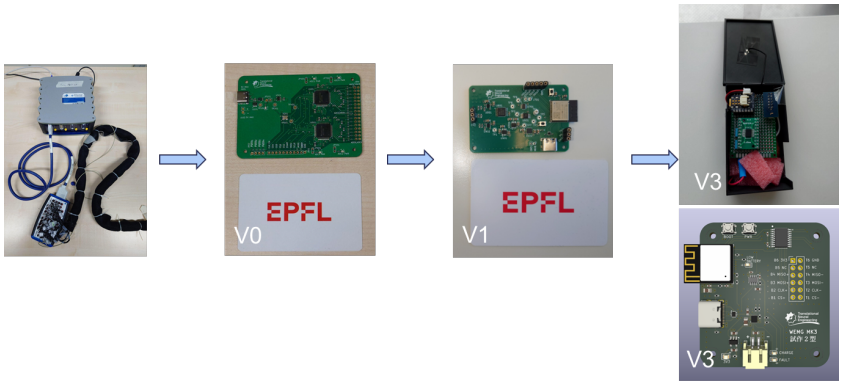


Figure 11.4: Migration from a commercial EEG acquisition system to prototypes of portable EMG acquisition devices. (Left) A clinical EEG acquisition system from gtec. (Second from the Left) EMG V0, a 32-channel wired custom EMG acquisition board, designed for idea evaluation. (Second from the Right) EMG V1, a 2-channel wireless EMG acquisition board, using time-multiplexed sampling architecture. (Right) EMG V3, a standalone 32-channel wireless EMG acquisition board, used in the final stage of the project.

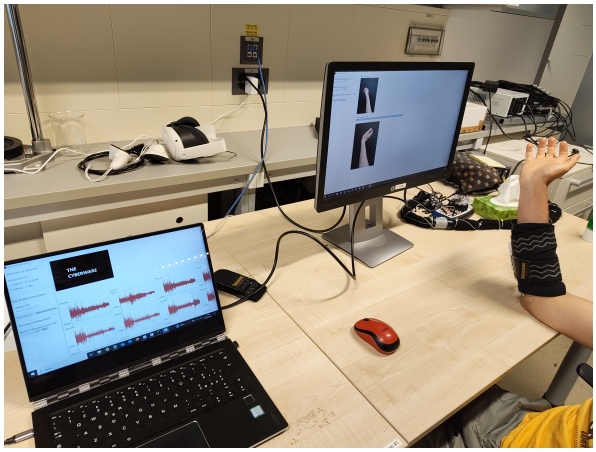
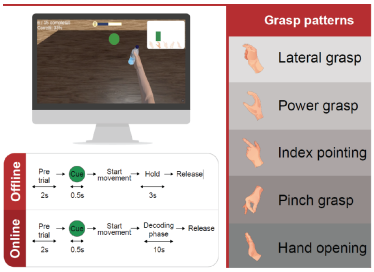


Figure 11.5: A subject wearing EMG V3 while following a hand movement guide. The subject's EMG signals were displayed in realtime on the computer on the left.

tbh]



Participants (offline exp)

ID	Sex	Age	Injury level	AIS	MRC finger flex.	MAS finger flex.
P1	**m**	**57**	**C3**	**D**	**2**	**0**
P2	m	48	C4	C	3	2
P3	m	81	C4	D	4	0
P4	m	46	C4	C	2	0
P5	m	49	C4	D	4	0
P6	**m**	**68**	**C4**	**D**	**2**	**0**

Controls

ID	Sex	Age
C1	f	30
C2	m	31
C3	m	28
C4	m	29

Figure 11.6: Experimental protocol and demographics of the patients and subjects recruited. Figure adapted from [Ceradini et al. 2024], with permission. Original poster presented at the Society for Neuroscience Annual Meeting, Chicago, United States.

comfort. The V3 acquisition system has been tested on more than 10 healthy subjects (Figure 11.5) in classifier-based (5-hand pose) intention decoding.

11.2.3 Decoding Improvements

To further improve real-time decoding from forearm EMG signals, our team focused on enhancing the precision of the calibration process and integrating new approaches for more reliable kinematic detection. Previously, finger kinematics were recorded using only two RGB cameras, with finger angles extracted using MediaPipe. However, MediaPipe's limitations in precision motivated us to explore alternative calibration strategies. To address this, we have integrated Mocap gloves, which provide far more accurate kinematic data and greatly improve the calibration process for EMG signals. This has been successfully integrated with our real-time training procedure, allowing for more reliable and precise EMG decoding.

In parallel, efforts have been made to reduce the computational cost of the model, with a focus on making it portable and suitable for lightweight, embedded systems. A new lightweight model was designed, significantly reducing the computational load while retaining high performance. The model, which includes an attractor-based neural network architecture with an encoder, attention layer, attractor network, and refinement regressor, achieved improved performance while reducing the number of parameters by 50 times compared to traditional models.

Accuracy wise, this new model achieved comparable performance compared to LSTM [Avian et al. 2022] and deep CNN [Mendez et al. 2021] while being more than 50 times smaller. The reduced model was tested on a Raspberry Pi 5 combined with a Google Coral TPU, which allowed us to further reduce the system's setup size and make it feasible for portable applications. The combination of the lightweight model

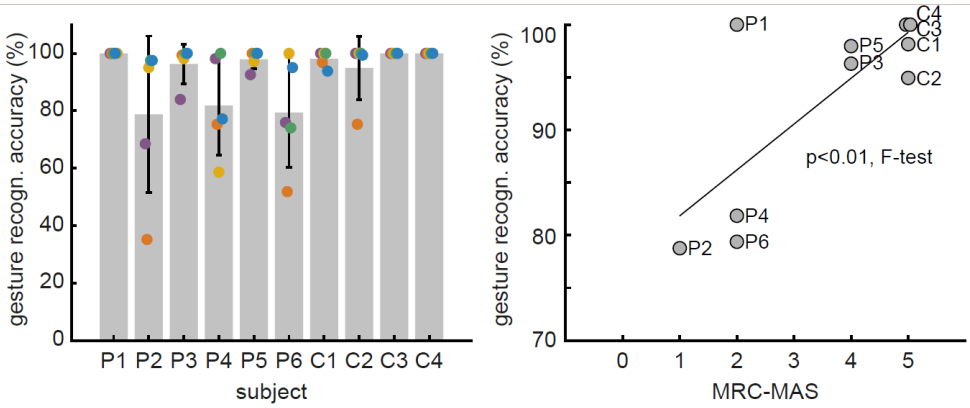


Figure 11.7: (Left) Offline classification accuracy on patients and healthy subjects. (Right) correlation between the level of impairment of the patients and offline performance. Figure adapted from [Ceradini et al. 2024], with permission. Original poster presented at the Society for Neuroscience Annual Meeting, Chicago, United States.

and edge hardware demonstrated the system's capability to perform efficient decoding while requiring minimal computational resources, making it ideal for wearable use. This model, which was introduced in our recent conference paper at AICAS 2024 [Kalbasi et al. 2024], offers an ideal balance of decoding accuracy, portability, and low power consumption.

In summary, by incorporating kinematic gloves for enhanced calibration and reducing the model size, we have significantly improved the system's real-time decoding capabilities. These enhancements are paving the way for practical applications of EMG-based control systems, particularly in the context of robotic prosthetic hands and other assistive technologies aimed at restoring motor function for people with disabilities.

Gesture Decoding with Spinal Cord Injury Patients

In our efforts to evaluate gesture decoding in individuals with spinal cord injuries (SCI), we conducted grasp classification tests involving six SCI patients and four healthy subjects. Each participant completed 60 repetitions of 5 different movements, allowing us to assess the system's robustness across varying conditions (Figure 11.6). The offline results demonstrated high decoding performance, achieving an accuracy of approximately 90% in patients.

This high level of decoding performance suggests that our system is capable of effectively interpreting EMG signals even in individuals with significant motor impairments. Moreover, the results showed that the decoding accuracy was correlated with

the functional level of each participant, which indicates that the system can adapt to different degrees of motor impairment (Figure 11.7). This adaptability is crucial for tailoring assistive technologies to meet the needs of individual users, allowing for a more personalized approach in rehabilitation or daily assistive devices.

The findings of this offline study provide a foundation for the development of the system, particularly its potential application aimed at improving the quality of life of individuals with SCI.

11.3 Conclusions

On the way to achieve highly accurate and robust user intention decoding for navigating in VR, we developed and optimized our decoding system both in software and hardware. Our non-invasive decoding system, after continuous improvements, have evolved into a wearable form factor, with greatly reduced noise levels. The system enables real-time, intuitive control of digital interfaces, prosthetics, and robotic systems. On the software side, as part of the decoding algorithm validation, we achieved high decoding accuracy (approximately 90%) in subjects with SCI in offline decoding, showing that our system can be used to "reconnect" people with motor disorders with digital interfaces to promote social interaction. We imagine the future of our system to be in an even smaller and multi-functional form. Therefore, future development could further miniaturize the hardware footprint and integrate the currently external decoder machine. Already within the SUN project, the technology was introduced and validated in the Case Study "Extended Reality for People with Serious Mobility and Verbal Communication Diseases" (Chapter 22).

References

Avian, Cries, Setya Widyawan Prakosa, Muhamad Faisal, and Jenq-Shiou Leu (2022). "Estimating finger joint angles on surface EMG using Manifold Learning and Long Short-Term Memory with Attention mechanism". In: *Biomedical Signal Processing and Control* 71, p. 103099.

Brucker, Bernard S. and Natalya V. Buylaeva (1996). "Biofeedback effect on electromyography responses in patients with spinal cord injury". In: *Archives of Physical Medicine and Rehabilitation* 77.2, pp. 133–137.

Campanini, Isabella, Catherine Disselhorst-Klug, William Z. Rymer, and Roberto Merletti (2020). "Surface EMG in Clinical Assessment and Neurorehabilitation: Barriers Limiting Its Use". In: *Frontiers in Neurology* Volume 11 - 2020.

Ceradini, M., E. Losanno, F. I. Serdana, V. Mendez, C. Chestek, G. Righi, G. Del Popolo, S. Shokur, and S. Micera (Oct. 2024). *Online myoelectric gesture recognition in patients with tetraplegia during attempted movements*. Abstract presented at the Society for Neuroscience Annual Meeting (SfN 2024). Abstract code: PSTR288.05/I10. Chicago, IL, United States.

Dwivedi, Anany, Yongje Kwon, and Minas Liarokapis (2020). "EMG-Based Decoding of Manipulation Motions in Virtual Reality: Towards Immersive Interfaces". In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3296–3303.

Englehart, Kevin B. and Bernard Hudgins (2003). "A robust, real-time control scheme for multifunction myoelectric control". In: *IEEE Transactions on Biomedical Engineering* 50, pp. 848–854.

Kalbasi, Mohammad, MohammadAli Shaeri, Vincent Alexandre Mendez, Solaiman Shokur, Silvestro Micera, and Mahsa Shoaran (2024). "A Hardware-Efficient EMG Decoder with an Attractor-based Neural Network for Next-Generation Hand Prostheses". In: *2024 IEEE 6th International Conference on AI Circuits and Systems (AICAS)*, pp. 532–536.

McManus, Lara, Giuseppe De Vito, and Madeleine M. Lowery (2020). "Analysis and Biophysics of Surface EMG for Physiotherapists and Kinesiologists: Toward a Common Language With Rehabilitation Engineers". In: *Frontiers in Neurology* Volume 11 - 2020.

Mendez, V., L. Pollina, F. Artoni, and S. Micera (2021). "Deep Learning with Convolutional Neural Network for Proportional Control of Finger Movements from surface EMG Recordings". In: *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1074–1078.

Nam, Chingyi, Wei Rong, Waiming Li, Chingyee Cheung, Wingkit Ngai, Tszching Cheung, Mankit Pang, Li Li, Junyan Hu, Honwah Wai, and Xiaoling Hu (2022). "An exoneuromusculoskeleton for self-help upper limb rehabilitation after stroke". In: *Soft robotics* 9, pp. 14–35.

Pan, Li-Ling Hope, Wen-Wen Yang, Chung-Lan Kao, Mei-Wun Tsai, Shun-Hwa Wei, Felipe Fregni, Vincent Chiun-Fan Chen, and Li-Wei Chou (2018). "Effects of 8-week sensory electrical stimulation combined with motor training on EEG-EMG coherence and motor function in individuals with stroke". In: *Scientific reports* 8, p. 9217.

Zhuang, Katie Z., Nicolas Sommer, Vincent Mendez, Saurav Aryan, Emanuele Formento, Edoardo D'Anna, Fiorenzo Artoni, Francesco Maria Petrini, Giuseppe Granata, Giovanni Cannaviello, Wassim Raffoul, Aude G Billard, and Silvestro Micera (2019).

"Shared human–robot proportional control of a dexterous myoelectric prosthesis". In: *Nature Machine Intelligence* 1, pp. 400–411.

12. Postural Assessment and Monitoring of Body Kinematics

Panagiotis Kasnesis[1], *Theodora Plavoukou*[1,2], *Lazaros Toumanidis*[1], *Amalia Contiero Syropoulou*[1], *and George Georgoudis*[1,2]

[1] ThinGenious PC (THING), Greece
[2] University of West Attica, Greece

Abstract. The wearable-based postural assessment component processes multimodal data from eight combined surface Electromyography (sEMG) and Inertial Measurement Unit (IMU) sensors to classify and evaluate lower-limb rehabilitation exercises in real time. A hierarchical deep learning model was developed to detect common rehabilitation exercises (squat, knee extension, and gait) and to distinguish correct from incorrect execution by the wearer. Data segmentation, normalization, and convolutional architectures were employed. Evaluation on a public dataset (KneE-PAD) achieved an accuracy of 83.75%, while fine-tuning with additional healthy-subject data improved accuracy to 91.3%. To support deployment in the SUN XR project, a real-time software application was developed comprising two components, Collect and Posture. The system enables live data collection, preprocessing, model inference, and feedback delivery, supporting continuous and automated rehabilitation exercise assessment in the Extended Reality (XR) environment.

12.1 Introduction

Lower-limb rehabilitation is an essential part of recovery after injury and surgery. It is performed through a series of structured and intermittent exercises aimed at rebuilding motor function, muscle strength, and joint stability. Accurate assessment of such exercises is crucial in order to avoid potential injury that might hinder the healing process.

Rehabilitation exercises are guided and supervised by clinicians, while correctness is assessed by visual inspection of the patient. Following in-person sessions in a hospital or rehabilitation center, patients are prescribed similar activities to continue their therapy at home for around three months, after which they will be reevaluated [Rejeski et al. 1997]. This part of the process is not supervised by a clinician; therefore, patients need to remain consistent to achieve their benefits [Jack et al. 2010]. To achieve this, virtual assistants have been suggested to remotely monitor and encourage patients via applications for Extended Reality (XR) [Sarri et al. 2024]. The majority of these use Machine Learning (ML)-based virtual coaches that can evaluate a patient's posture while they are performing the exercises [Biebl et al. 2020].

Recent advances in wearable sensing technologies have emerged, allowing for new possibilities for objective and continuous assessment of the recovery process [García-de-Villa et al. 2022; Nishiwaki et al. 2006; Diraneyya et al. 2021]. Surface electromyography (sEMG) can track muscle activation patterns [Nishiwaki et al. 2006], while inertial measurement units (IMUs) provide further information about limb orientation, acceleration, and angular velocity [García-de-Villa et al. 2022]. Using this data with deep learning methods may provide a deeper insight into the execution of exercises. However, this opportunity doesn't come without its challenges. Raw sEMG signals can vary highly across individuals and are prone to signal noise and artifacts [Truong et al. 2023]. IMU signals, while more stable, are subject to drift over time, which can complicate assessment on exercises with long duration [Trumble et al. 2017].

To address these challenges, we developed a wearable-based postural assessment component that integrates eight sEMG and IMU channels to classify and evaluate commonly prescribed lower-limb rehabilitation exercises.

The developed technology has been implemented and validated in the SUN project Case Study: "Extended Reality for Rehabilitation" (Chapter 20)".

12.2 Methodology and Results

12.2.1 Data Acquisition and Preprocessing

This experiment was conducted utilizing the KneE-PAD (Knee Rehabilitation Exercises for Postural Assessment Dataset) [Kasnesis et al. 2024], which we previously collected and published. KneE-PAD contains sEMG and IMU recordings of lower-limb rehabilitation exercises, performed by patients who experienced clinically diagnosed knee pathologies. The data were acquired from 31 participants, ranging from 18-68 years

old in age, 152-200 cm in height, and 57-146 kg in weight. They executed three rehabilitation exercises, which are squat, seated leg extensions, and simple gait, without being provided direct supervision, thus emulating home-based rehabilitation conditions. For each exercise, two common incorrect patterns were identified and labeled, yielding nine distinct classes: three correct and six incorrect performance variants, described below:

- *Squat (Correct)*: From a standing position, the participant performs a seated motion toward a chair, avoiding lateral deviation of the knees or hips, and returns to standing upon lightly touching the chair;
- *Squat_WT (Wrong)*: Squat performed with body weight shifted predominantly onto the healthy leg;
- *Squat_FL (Wrong)*: Squat performed while positioning the injured leg forward;
- *Extension (Correct)*: While seated, starting from 90° knee flexion, the healthy leg remains stationary while the injured leg is extended in the sagittal plane until reaching full extension;
- *Extension_NF (Wrong)*: Knee extension exercise performed without achieving the full range of motion;
- *Extension_LL (Wrong)*: Knee extension performed while lifting the leg off the chair;
- *Gait (Correct)*: The participant rises from a chair (45–50 cm high, no armrests), stands upright, walks freely for 3 m, turns, and returns to the starting position;
- *Gait_NF (Wrong)*: Walking performed without fully extending the knee of the injured leg;
- *Gait_HA (Wrong)*: Walking performed with the injured leg fully extended at the knee while abducting the hip.

In particular, a total of 2,086 files were collected, with each one having an approximate duration of 4.2s and containing around 87.75M sEMG and 61.94M IMU samples.

It should be noted that this study was approved by the Research Ethics Committee of the University of West Attica (Approval No. 65417, 29/07/2023). Furthermore, all participants provided written informed consent for their data to be collected, processed, and published in accordance with the relevant regulations and the principles of the Declaration of Helsinki.

As for the sensor configuration, eight Delsys Trigno Avanti wearable sensors [Delsys avanti 2023] were employed. Each device integrates an sEMG module and a six-axis

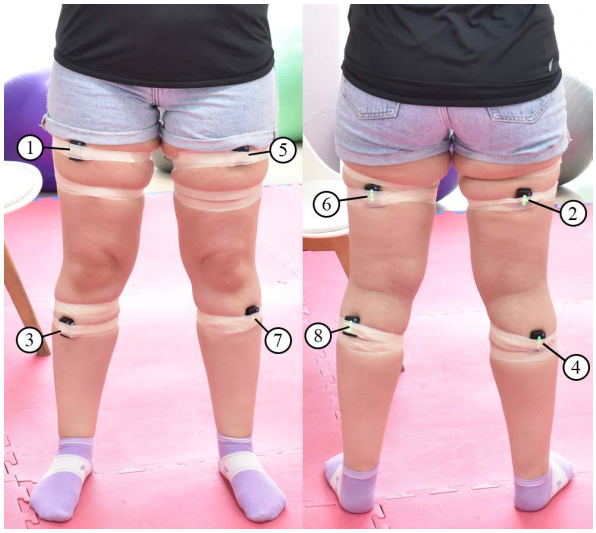


Figure 12.1: Placement of the eight Delsys Trigno Avanti sensors: 1–4 on the right leg (rectus femoris, hamstrings, tibialis anterior, gastrocnemius) and 5–8 on the left leg in the same order. Figure adapted from [Kasnesis et al. 2025] licensed under CC-BY 4.0.

IMU comprising a triaxial accelerometer as well as a triaxial gyroscope. These sensors are compact and lightweight, each sized 27 $\times$ 37 $\times$ 13 mm with a weight of 14 g, while their battery can last up to 8 hours of use. On each leg, four sensors were placed, two in front and two in the back, as illustrated in Figure 12.1. The muscle sites of interest were the rectus femoris, hamstrings, tibialis anterior, and gastrocnemius. Moreover, we configured the sEMG sensor at 1259.259 Hz as the sampling rate and the IMU sensor at 148.148 Hz. The sEMG range is 11 millivolt (mV), having a bandwidth of 20-450 Hz, the accelerometer range is $\pm$ 2 g at a 24–470 Hz bandwidth and the gyroscope's range is 250 deg/s with a bandwidth of 24-360 Hz.

Before being used in the deep learning model, the raw sEMG signals were processed to reduce motion artifacts and baseline drift by applying a second-order Butterworth high-pass filter with a 50 Hz cut-off. Values exceeding 0.3 mV were clipped as artifacts. IMU data were synchronized with the processed sEMG channels to form time-aligned multimodal input. All recordings were segmented into 4-second windows to standardize input length for deep learning. In addition, a sliding-window approach was used, with a step size of 250 ms for static exercises (squats and extensions) and 500 ms for gait, thus balancing temporal continuity with data augmentation.

The dataset, after segmentation, consisted of 4833 labeled samples (Squat: 840, Squat_WT: 475, Squat_FL: 432, Extension: 404, Extension_NF: 318, Exten-

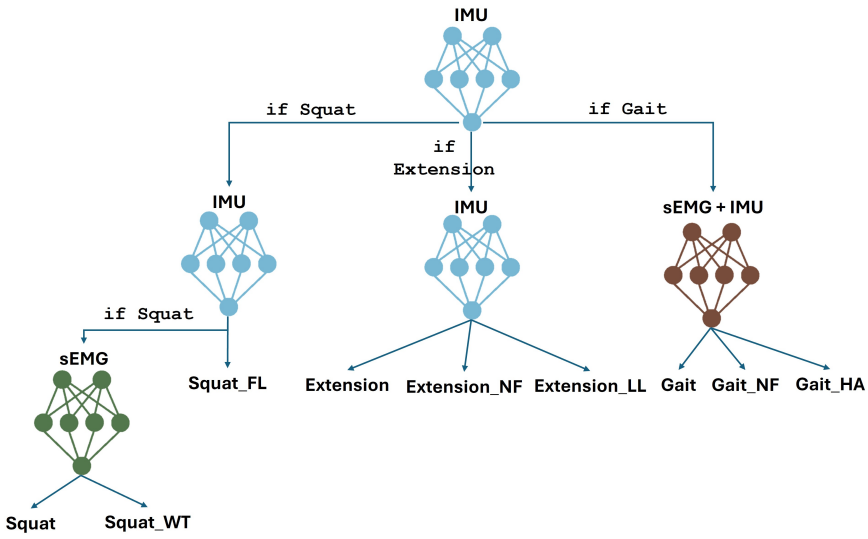


Figure 12.2: The proposed HDL architecture for rehabilitation exercise assessment.

sion_LL: 480, Gait: 485, Gait_NF: 704, Gait_HA: 695). Each segment resulted in an $8 \times N$ matrix for the sEMG data, and a $48 \times M$ matrix for the IMU data, where N and M represent the sampling rates and window lengths, respectively. All files were saved in NumPy (`.npy`) format, in order to be easily loaded and preprocessed within the training pipeline.

12.2.2 Model Architecture

Utilizing existing research [Villa et al. 2024] and some experimentation, we developed a Hierarchical Deep Learning (HDL) system for classifying and interpreting rehabilitation exercises collected from multimodal wearable data. The HDL system is comprised of five convolutional subnetworks connected in a hierarchy that first identifies which exercise was performed and then evaluates whether it was performed correctly or incorrectly.

This HDL system consists of five convolutional subnetworks arranged in a hierarchical structure. The top level is a Human Activity Recognition (HAR) network that inputs the 48-channel IMU input and outputs a coarse classification between squat, extension, or gait. This determination routes the example into one of three exercise-specific branches, which are each specialized in evaluating the correctness of the assessment. For gait assessment, both sEMG and IMU signals are used. For extension assessment, only IMU channels are employed. For squat assessment, the branch is further divided:

if the HAR output is squat, an IMU-based subnetwork first differentiates between a correct squat and the Squat_FL variation. If classified as not Squat_FL, a second subnetwork analyzes the sEMG channels alone to distinguish correct squats from the Squat_WT variation. The overall hierarchical structure of the proposed HDL system is illustrated in Figure 12.2.

Each of the five subnetworks has a similar architecture based on the convolutional feature extraction. Each has 1D convolutional layers with a ReLU activation function, max pooling and a dropout for normalization. Specifically, four convolutional blocks are used to process the IMU inputs, five are used for sEMG inputs due to more complex and noisier patterns of muscle activations in sEMG inputs. When multimodal input (IMU and sEMG) is required as input, the features were extracted separately and then fused together through modality-wise convolution before routing to the dense layers. The final classification layers are fully connected, using softmax outputs for the respective class. It's important to note that the subnetworks do not share weights, allowing each of them to learn task-specific features. This hierarchical structure allows for the model to use specialized networks for each exercise type while maintaining low latency and high classification performance.

12.2.3 Training and Evaluation

All subnetworks of the proposed Hierarchical Deep Learning (HDL) architecture were implemented and trained using supervised learning on the preprocessed KneE-PAD segments. To identify optimal configurations, a random hyperparameter search was conducted over convolutional kernel sizes, dropout rates, and learning rates prior to the final training, using the Leave-One-Session-Out (LOSO) cross-validation (CV) technique. Thirty subjects were used as the training set and one subject was used as the test set, repeating this process 31 times so that every subject was examined as a test set. With this method, we validated the model's generalizability to unseen subject data.

The Adam optimizer [Kingma and Ba 2015] was used with an initial learning rate of 0.001. A batch size of 256 and a maximum of 100 epochs were employed, with early stopping applied using a patience of 25 epochs to prevent overfitting. The weights of the validation model that achieved the best accuracy were stored and used for evaluation on the test set.

12.2.4 Results

The proposed HDL network achieved an accuracy of 83.75% and an F1-score of 80.64%. The confusion matrix of the HDL network is shown in Figure 12.3. It is

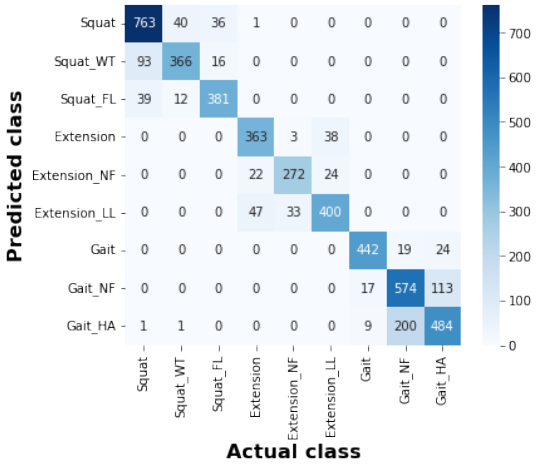


Figure 12.3: Confusion matrix of the proposed HDL network.

important to note that if the HAR network outputs an incorrect label, then the sample is passed to an incorrect branch, as the true labels are not known during inference. The confusion matrix shows that only three misclassifications occurred. This is beneficial because the subsequent subnetworks evaluate one exercise type at a time rather than all types concurrently, which would otherwise result in lower overall performance (accuracy of 73.75% and an F1-score of 65.82%). Lastly, the greatest classification confusion occurs between Gait_NF and Gait_HA, which is less critical if all gait-related errors are bundled as a single class.

12.2.5 Extra Evaluation

To further improve upon the aforementioned results, we created an additional evaluation dataset consisting of three healthy participants, who completed 10 repetitions of each exercise. When testing the original trained model on this new dataset, it achieved just 48% accuracy. This was impacted not only by the algorithm's sensitivity, but also by the fact that, by utilizing the LOSO technique, one of the 31 trained model versions must be chosen. Thus, having a dedicated dataset for selecting the best-performing version was deemed essential.

We fine-tuned the HDL system, which consists of the five subnetworks discussed in Subsection 12.2.2, while employing the same classification pipeline. After fine-tuning on each subnetwork, we then used the new dataset to find the best model version. After this process, the new HDL model achieved an accuracy of 91.3% on the new

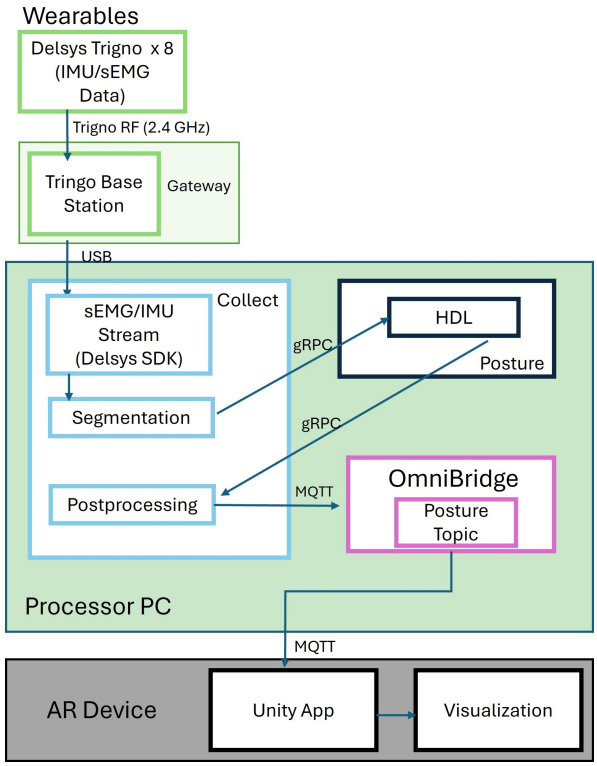


Figure 12.4: The developed pipeline of services consisting of the *Collect* and the *Posture* project. The former includes the data streaming module, the data preprocessing (segmentation), the data postprocessing, and sends the messages to the OmniBrige. The latter consists of the normalization function and the HDL network.

dataset, while maintaining the same performance on the KneE-PAD dataset, with only a 0.83% drop in F1-score.

12.2.6 Software Application

A real-time postural assessment application was developed to evaluate the HDL model in the Case Studies implementation of the presented methods. In particular, the developed desktop application applies the whole data (post/pre)processing pipeline and communicates with the MQTT broker. It is fully dockerized and documented[1], and is

[1]https://collect-thingenious-sun-75dfc78b8e799dad8a229fbab2ca44b424d055c.gitlab.io/

uploaded to SUN's GitLab repository[2][3]. Moreover, we used NVIDIA Triton inference server[4] to deploy the HDL model that supports dynamic batching and streaming.

The developed pipeline is described in an abstract manner in Figure 12.4. As shown, the app consists of two Docker apps, the *Collect* and the *Posture*. A script initializes them using *docker run*. Afterwards, the *Collect* app connects to the base station using Delsys SDK and waits until it is triggered by the AR device. When a start message is received on the corresponding MQTT topic, it starts collecting data from sensors, while when a stop message is received, it stops the data collection process. Afterwards, when a batch of data is collected (4 secs), it sends it via gRPC protocol to get a prediction from the HDL model, processes it, and publishes the prediction to a topic in OmniBridge.

12.3 Conclusion

In this work, we presented a wearable-based postural assessment component designed to objectively classify and evaluate lower-limb rehabilitation exercises in real time. By integrating multimodal data from eight sEMG and IMU sensors, we developed an HDL system that leverages specialized convolutional subnetworks to first identify the performed exercise and then assess its correctness.

The publicly available KneE-PAD dataset was used for extensive experimental evaluation of the HDL architecture. We achieved an accuracy of 83.75% and an F1-score of 80.64%. Most errors were made within the same variants of exercises. Subsequent fine-tuning with additional healthy-subject data further increased accuracy to 91.3% while maintaining generalizability. We also implemented a real-time software system comprising of two components, Collect and Posture, enabling live data collection, preprocessing, and inference.

These findings demonstrate the potential of multimodal sensing and hierarchical modeling to provide reliable, automated feedback during home-based rehabilitation, minimizing the need for ongoing clinical oversight and supporting safe exercise performance.

[2] https://evia.in-two.com/sun/posture_data_collection
[3] https://evia.in-two.com/sun/posture_app
[4] https://developer.nvidia.com/triton-inference-server

References

Biebl, Johanna Theresia, Marzena Rykala, Maximilian Strobel, Pawandeep Kaur Bollinger, Bernhard Ulm, Eduard Kraft, Stephan Huber, and Andreas Lorenz (2020). "App-Based Feedback for Rehabilitation Exercise Correction in Patients With Knee or Hip Osteoarthritis: Prospective Cohort Study". In: *Journal of Medical Internet Research* 23.

Delsys avanti (2023). *Trigno Avanti Sensor Superior EMG + IMU Technology*. https://delsys.com/trigno-avanti/.

Diraneyya, Mohsen Mutasem, JuHyeong Ryu, Eihab M. Abdel-Rahman, and Carl T. Haas (2021). "Inertial Motion Capture-Based Whole-Body Inverse Dynamics". In: *Sensors (Basel, Switzerland)* 21.

García-de-Villa, Sara, Ana Jiménez-Martín, and Juan Jesús García-Domínguez (2022). "A database of physical therapy exercises with variability of execution collected by wearable sensors". In: *Scientific Data* 9.

Jack, Kirsten, Sionnadh McLean, Jennifer A Klaber Moffett, and Eric Gardiner (2010). "Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review". In: *Manual Therapy* 15, pp. 220–228.

Kasnesis, Panagiotis, Theodora Plavoukou, Amalia Contiero Syropoulou, George Georgoudis, and Lazaros Toumanidis (2024). *KneE-PAD [Data set]*. Zenodo https://doi.org/10.5281/zenodo.12112951.

Kasnesis, Panagiotis, Theodora Plavoukou, Amalia Contiero Syropoulou, Lazaros Toumanidis, and George Georgoudis (2025). "A Knee Rehabilitation Exercises Dataset for Postural Assessment using Wearable Devices". In: *Scientific Data* 12.1, p. 610.

Kingma, D. P. and J. Ba (2015). "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980.

Nishiwaki, Gaston Ariel, Yukio Urabe, and Kosuke Tanaka (2006). "EMG Analysis of Lower Extremity Muscles in Three Different Squat Exercises." In: *Journal of the Japanese Physical Therapy Association = Rigaku ryoho* 9 1, pp. 21–6.

Rejeski, Walter Jack, Lawrence R Brawley, Walter H. Ettinger, Timothy R. Morgan, and Christopher Thompson (1997). "Compliance to exercise therapy in older participants with knee osteoarthritis: implications for treating disability." In: *Medicine and science in sports and exercise* 29 8, pp. 977–85.

Sarri, Francesca, P. Kasnesis, S. Symeonidis, I. Th, Paraskevopoulos, S. Diplaris, F. Posteraro, Georgios Georgoudis, and K Mania (2024). "Embodied Augmented Reality for Lower Limb Rehabilitation". In.

Trumble, Matthew, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse (2017). "Total capture: 3d human pose estimation fusing video and inertial sensors". In: *Proceedings of 28th British Machine Vision Conference*, pp. 1–13.

Truong, Minh Tat Nhat, Amged Elsheikh Abdelgadir Ali, Dai Owaki, and Mitsuhiro Hayashibe (2023). "EMG-Based Estimation of Lower Limb Joint Angles and Moments Using Long Short-Term Memory Network". In: *Sensors (Basel, Switzerland)* 23.

Villa, Sara García de, David Casillas-Pérez, Ana Jiménez-Martín, and Juan Jesús García Domínguez (2024). "Simultaneous exercise recognition and evaluation in prescribed routines: Approach to virtual coaches". In: *ArXiv* abs/2401.12857.

13. Multimodal Pose Estimation

Vasileios-Rafail Xefteris[1], *Amalia Contiero Syropoulou*[2], *Panagiotis Kasnesis*[2], *Spyridon Symeonidis*[1], *Sotiris Diplaris*[1], *and Stefanos Vrochidis*[1]

[1] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece
[2] ThinGenious PC (THING), Greece

Abstract. This chapter describes the multimodal 3D pose estimation component developed in the context of the SUN project. The multimodal 3D pose estimation component is based on the development and fusion of two different unimodal components, the sensor-based and the visual-based 3D pose estimation models. The sensor-based component is based on a set of six Inertial Measurement Unit (IMU) sensors using the TIP (Transformer Inertial Poser) architecture. The computer vision model was developed by adopting MediaPipe's 3D Body Landmarker to analyze color video recordings and extract 3D human joint positions. The MediaPipe's output was enriched in order to also extract bone rotations, using a neural network approach. Finally, the fusion of the sensor-based and vision-based model outputs is based on a hybrid LSTM-Random Forest network to capture the temporal characteristics of body poses. The multimodal 3D pose estimation and the vision-based 3D pose estimation components have been deployed in the SUN project lower limb and upper limb rehabilitation scenario, respectively.

13.1 Introduction

The capturing and monitoring of human motions are of increasing interest in a variety of applications, including, among others, Virtual Reality (VR) [Mehta et al. 2017], Human-Computer Interaction (HCI) [Huo et al. 2023], entertainment [Zhang et al. 2015] and healthcare [Xu et al. 2022]. The recent advancements in computer vision and inertial sensing enable effective human pose estimation in less controlled environments,

eliminating setup constraints and reducing the burden of specialized equipment that restricts free movement.

Concerning computer vision methods, modern pipelines typically detect 2D joint keypoints with high-resolution CNNs (e.g., OpenPose [Martınez 2019]) and then "lift" those detections into 3D using either temporal models (e.g., VideoPose3D [Pavllo et al. 2019]) or graph-based networks (e.g., GraphCMR [Kolotouros et al. 2019]). Real-time single-RGB approaches like VNect demonstrated early interactive performance, while on-device frameworks such as Google's MediaPipe Body Landmarker[1] now offer lightweight, efficient 3D inference for both desktop and mobile platforms.

In parallel, wearable-based pose estimation has advanced significantly. Early approaches such as the Sparse Inertial Poser (SIP) [Marcard et al. 2017] demonstrated that full-body joint angles can be reconstructed from as few as six Inertial Measurement Units (IMUs) by optimizing recorded signals offline without motion priors. Subsequent methods, such as Deep Inertial Poser (DIP) [Huang et al. 2018], used bidirectional recurrent networks that were trained on huge synthetic datasets like AMASS [Mahmood et al. 2019], and were able to provide low-latency online estimation. Recent methods like TransPose [Yi et al. 2021], Physical Inertial Poser [Yi et al. 2022], and Transformer Inertial Poser (TIP) [Jiang et al. 2022], utilized Transformer-based architectures and incorporated sensible physical constraints, leveraging kinematics or analytical methods [Jiang et al. 2022], to reduce drift and provide low mean joint errors while still maintaining sufficient real-time performance. These developments have led to sparse IMU configurations being a lightweight alternative compared to more complicated multi-sensor methods.

Both computer vision and inertial sensor-based solutions face difficulties that are based on their characteristics. Vision-based solutions demand a complex human model [Von Marcard et al. 2017] and are prone to occlusions. On the other hand, inertial sensor approaches lack positional information and suffer from drift over even short time periods. The fusion of the two modalities can overcome these limitations by effectively and robustly combining visual-based and inertial sensor-based results, ensuring high levels of accuracy in real-time human pose estimation. Works like [Trumble et al. 2017] and [Huang et al. 2020] proposed methods based on deep learning networks, which can learn the unique characteristics of each modality.

In the context of the SUN project, the multimodal 3D pose estimation is a key component of the lower limb rehabilitation scenario. The component is responsible for capturing and estimating the patient's joint movements in real-time and forwarding the estimations to the avatar design component.

[1] https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker

13.2 Methodology and Results

The current section presents detailed methods and results for the unimodal sensor-based and vision-based pose estimation components, as well as the multimodal fusion component, in the context of 3D human pose estimation.

13.2.1 Sensor-Based Pose Estimation

The sensor-based approach utilizes six IMUs affixed to the participant's head, waist and four limb positions. In this study, the adopted framework is TIP [Jiang et al. 2022]. This system integrates a learning-driven Transformer-decoder architecture [Radford et al. 2018] with an analytical drift-compensation mechanism. The Transformer network infers full-body joint angles, three-dimensional coordinates, root velocity and Stationary Body Points (SBPs) from the orientation and acceleration signals provided by the IMUs. Concurrently, the drift-stabilization module refines these output online by leveraging SBPs to enhance motion fidelity and mitigate cumulative error over time.

TIP is trained on selected subsets of the AMASS repository [Mahmood et al. 2019], adapted to the SMPL-H model [Loper et al. 2023]. As AMASS does not include genuine IMU recordings, synthetic inertial data are generated following the procedure outlined in DIP [Huang et al. 2018]. The model receives as input calibrated orientation and acceleration streams, processed through smoothing and integration, collected from IMUs placed at the waist, wrists, knees and head. The network predicts 18 SMPL-defined joint angles (excluding fingers, wrists, and toes), the root's linear velocity, and SBP coordinates. A temporal window of size 39 frames is employed, with the root orientation derived directly from the waist IMU.

The Transformer Motion Estimator (TME) is implemented as a conditional Transformer decoder [Radford et al. 2018]. During training, the TME processes sequences of IMU measurements alongside its own previous predictions shifted by one timestep. These concatenated inputs are projected through a linear layer before being passed to four sequential Transformer blocks. Each block consists of a multi-head attention mechanism and a feedforward sub-network, connected through residual pathways and followed by layer normalization. The resulting embeddings are then summarized by a unidirectional recurrent layer with hyperbolic tangent activation. A final linear layer produces the estimated joint angles, root velocity, and SBPs. The network is trained with standard metrics, which include mean squared error for the joint angles, root velocity, and SBP coordinates, and binary cross-entropy for SBP activation flags. Predicted SBPs are iteratively fed back into the input history for subsequent steps.

Upon generation of these predictions, the drift-compensation module applies online corrections to both root motion and joint angles. SBP predictions serve as constraints to adjust root velocity, ensuring horizontal motion stability. For joint-angle refinement, when two SBPs remain consistently active across successive frames, a two-bone inverse kinematics (IK) [Holden et al. 2020] method is employed to preserve their spatial relationship, while respecting the TME's dependence on prior output. The corrected joint angles are then incorporated back into the historical buffer for future inference. It should be noted that this module employs a minimal version of TIP without its terrain-generation module. Since the TotalCapture dataset was used for evaluation, which does not involve uneven ground, this component is unnecessary and its omission does not affect the model's predictions, as it is intended solely for visualization purposes.

13.2.2 Vision-Based Pose Estimation

In the SUN project, vision-based pose estimation is used in the Case Study "Extended Reality for Rehabilitation" (Chapter 20) for both the upper and lower limb use cases. In order to address the specific needs of each use case, two modules were created: one targeting upper-limb movements and one for the lower-limb exercises. For the upper limb use case, the module calculates arms' and elbows' angles and checks the body's posture (e.g., if there is a torso leaning, if shoulders are higher than normal), providing feedback to the system to ensure correct posture of the patient during the execution of specific movements. The module implemented for the lower limb use case estimates the user's full-body 3D pose using RGB input in real-time, while it also maps the detected human body landmarks on 3D human avatars inside the Extended Reality (XR) environment. Its output is sent directly to the multimodal fusion component, where computer vision and IMU results are combined for the final mapping of user movements to a 3D avatar in the XR environment. For both vision-based modules, input is received from an external RGB camera, and MediaPipe's 3D Body Landmarker[2] is used.

It is worth noting that during the SUN project's research phase, a method to enrich MediaPipe's output was developed, which was presented in the IMX Workshop in Stockholm in June 2024 [Poulios et al. 2024]. The purpose of the proposed method is twofold. Firstly, to eliminate MediaPipe's pose estimation errors, primarily on depth estimation. Secondly, to enrich MediaPipe's output with extra information to include the longitudinal bones' rotations. To achieve both of our goals, we used the TotalCapture dataset to train two different neural networks, shown in Figure 13.1 and Figure 13.2.

[2] https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker

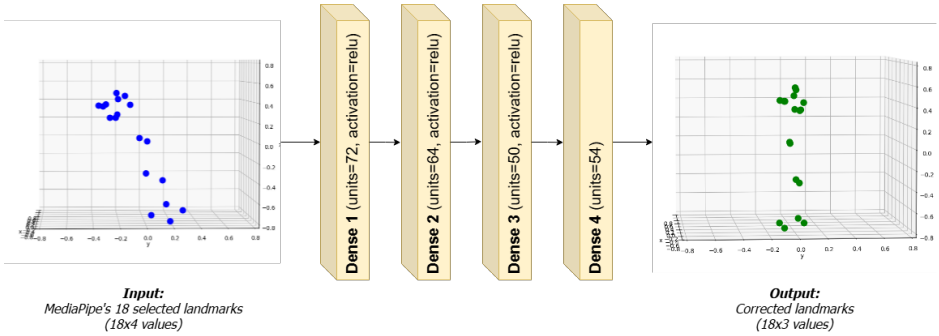


Figure 13.1: Landmarks' correction network architecture.

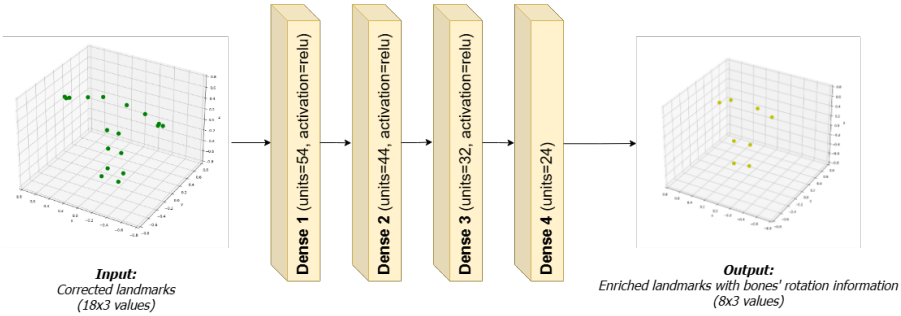


Figure 13.2: Bones' rotation information enrichment network architecture.

13.2.3 Multimodal Pose Estimation

As mentioned in the Section 13.1, the challenges that computer vision and inertial sensing solutions face can be addressed by their multimodal fusion. Therefore, a multimodal fusion module for pose estimation is designed in order to merge results from the visual-based and sensor-based pose estimation modules. MediaPipe's 3D landmark position estimations and TIP's 3D joint position predictions can serve as input to the developed multimodal fusion architecture. The multimodal fusion scheme is based on a decision-level fusion approach, where the results from both unimodal components are fused into a unified 3D joint position estimation.

The multimodal fusion module is based on a hybrid architecture, combining a deep learning network architecture with a machine learning regression model as the final output layer. Regarding the deep learning network architecture, a Long Short Term Memory (LSTM) based network was selected in order to exploit the temporal characteristics of the 3D human poses. The LSTM network consists of an LSTM cell with

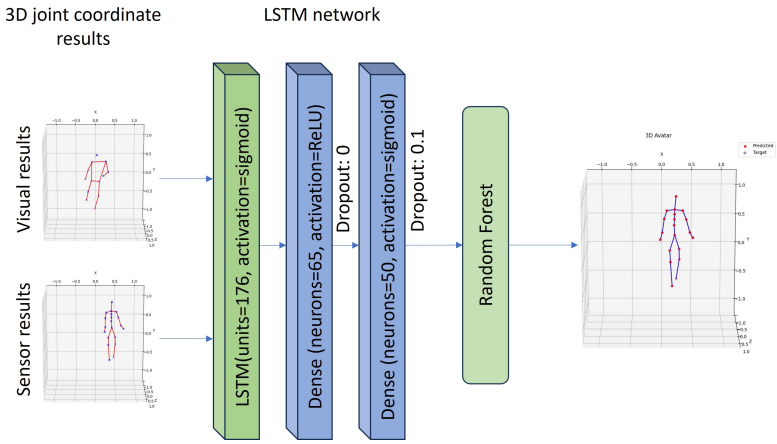


Figure 13.3: The hybrid LSTM-RF network architecture. Figure adapted from [Xefteris et al. 2025], licensed under CC-BY 4.0.

Table 13.1: Mean per joint position error results of the unimodal and multimodal components on the TotalCapture public dataset.

Method	MPJPE (cm)
Unimodal analysis results	
Visual sensors - MediaPipe	15.21
Inertial sensors - TIP	5.48
Multimodal analysis results	
LSTM-RF hybrid network	1.21

176 units and a look back of 5 frames, followed by two dense layers with 65 and 50 neurons, respectively.

The last layer of the hybrid network is a Random Forest (RF) regressor. The RF model is based on multiple decision trees with their individual outputs been averaged in order to produce the final model's output. The ensemble nature of RF is able to improve generalization and smooth out noise.

The full architecture of the hybrid LSTM-RF network can be seen in Figure 13.3. The results of the current configuration and the unimodal components on the TotalCapture public dataset can be seen in Table 13.1. From the Table, it can be seen that the multimodal fusion architecture improves the overall performance of the unimodal components by leveraging their unique characteristics, achieving a mean per joint position error of 1.21 cm. The presented methodology, including both the unimodal components and the hybrid LSTM-RF multimodal fusion, is also presented in [Xefteris et al. 2025].

13.3 Conclusions

The current chapter describes the whole multimodal 3D pose estimation process of the SUN project. The framework includes two unimodal components, namely the sensor-based component and the vision-based component, and a multimodal fusion component, responsible for fusing the outcomes of the unimodal components into a unified outcome. The sensor-based component employs six IMUs, utilizing the TIP architecture to estimate full-body joint angles, root velocity, and SBPs, with analytical drift compensation for improved stability of pose estimation over time. The computer vision-based component captures input from an RGB camera and adapts MediaPipe's 3D Body Landmarker to detect body landmarks and estimate their positions. The MediaPipe's outcome was further processed to eliminate errors due to camera positioning and depth estimation, and enriched to include also longitudinal bones' rotations. The multimodal fusion component is based on a hybrid machine learning and deep learning architecture, combining an LSTM-based network with a Random Forest regression algorithm as the final network layer, thereby exploiting the complementary strengths of both deep learning and machine learning models. The framework is designed to operate in a minimal setup, using a single camera and six IMU sensors. The designed system leverages the distinct but reinforcing aspects of inertial sensing and computer vision, thus improving the unimodal performance through the multimodal fusion of sensor and visual modalities.

References

Holden, Daniel, Oussama Kanoun, Maksym Perepichka, and Tiberiu Popa (2020). "Learned motion matching". In: *ACM Transactions on Graphics (TOG)* 39.4, pp. 53–1.

Huang, Fuyang, Ailing Zeng, Minhao Liu, Qiuxia Lai, and Qiang Xu (2020). "DeepFuse: An IMU-aware network for real-time 3D human pose estimation from multi-view image". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 429–438.

Huang, Yinghao, Manuel Kaufmann, Emre Aksan, Michael J. Black, Otmar Hilliges, and Gerard Pons-Moll (2018). "Deep inertial poser". In: *ACM Transactions on Graphics (TOG)* 37, pp. 1–15.

Huo, Rongtian, Qing Gao, Jing Qi, and Zhaojie Ju (2023). "3d human pose estimation in video for human-computer/robot interaction". In: *International Conference on Intelligent Robotics and Applications*. Springer, pp. 176–187.

Jiang, Yifeng, Yuting Ye, Deepak Edakkattil Gopinath, Jungdam Won, Alexander W. Winkler, and C. Karen Liu (2022). "Transformer Inertial Poser: Real-time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation". In: *SIGGRAPH Asia 2022 Conference Papers*.

Kolotouros, Nikos, Georgios Pavlakos, and Kostas Daniilidis (2019). "Convolutional mesh regression for single-image human shape reconstruction". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4501–4510.

Loper, Matthew, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black (2023). "SMPL: A skinned multi-person linear model". In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 851–866.

Mahmood, Naureen, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black (2019). "AMASS: Archive of motion capture as surface shapes". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5442–5451.

Marcard, Timo von, Bodo Rosenhahn, Michael J. Black, and Gerard Pons-Moll (2017). "Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs". In: *Computer Graphics Forum* 36.

Martınez, Ginés Hidalgo (2019). "Openpose: Whole-body pose estimation". In: *Ph. D. dissertation*.

Mehta, Dushyant, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt (2017). "Vnect: Real-time 3d human pose estimation with a single rgb camera". In: *Acm transactions on graphics (tog)* 36.4, pp. 1–14.

Pavllo, Dario, Christoph Feichtenhofer, David Grangier, and Michael Auli (2019). "3d human pose estimation in video with temporal convolutions and semi-supervised training". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7753–7762.

Poulios, Ilias, Theodora Pistola, Spyridon Symeonidis, Sotiris Diplaris, Konstantinos Ioannidis, Stefanos Vrochidis, and Ioannis Kompatsiaris (2024). "Enhanced real-time motion transfer to 3D avatars using RGB-based human 3D pose estimation". In: *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops*, pp. 88–99.

Radford, Alec, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. (2018). "Improving language understanding by generative pre-training". In.

Trumble, Matthew, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse (2017). "Total capture: 3d human pose estimation fusing video and inertial sensors". In: *Proceedings of 28th British Machine Vision Conference*, pp. 1–13.

Von Marcard, Timo, Bodo Rosenhahn, Michael J Black, and Gerard Pons-Moll (2017). "Sparse inertial poser: Automatic 3d human pose estimation from sparse imus". In: *Computer graphics forum*. Vol. 36. 2. Wiley Online Library, pp. 349–360.

Xefteris, Vasileios-Rafail, Amalia Contiero Syropoulou, Theodora Pistola, Panagiotis Kasnesis, Ilias Poulios, Athina Tsanousa, Spyridon Symeonidis, Sotiris Diplaris, Kostas Goulianas, Periklis Chatzimisios, et al. (2025). "Multimodal fusion of inertial sensors and single RGB camera data for 3D human pose estimation based on a hybrid LSTM-Random forest fusion network". In: *Internet of Things* 29, p. 101465.

Xu, Wei, Donghai Xiang, Guotai Wang, Ruisong Liao, Ming Shao, and Kang Li (2022). "Multiview Video-Based 3-D Pose Estimation of Patients in Computer-Assisted Rehabilitation Environment (CAREN)". In: *IEEE Transactions on Human-Machine Systems* 52.2, pp. 196–206.

Yi, Xinyu, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu (2022). "Physical Inertial Poser (PIP): Physics-aware Real-time Human Motion Tracking from Sparse Inertial Sensors". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13157–13168.

Yi, Xinyu, Yuxiao Zhou, and Feng Xu (2021). "TransPose: Real-Time 3D Human Translation and Pose Estimation with Six Inertial Sensors." In: *ACM Transactions on Graphics (TOG)* 40, pp. 1–13.

Zhang, Yizhai, Kuo Chen, Jingang Yi, Tao Liu, and Quan Pan (2015). "Whole-body pose estimation in human bicycle riding using a small set of wearable sensors". In: *IEEE/ASME Transactions on Mechatronics* 21.1, pp. 163–174.

14. Hand Gesture Recognition

Spyridon Symeonidis[1], *Sotiris Diplaris*[1], *and Stefanos Vrochidis*[1]

[1] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece

Abstract. With the increasing development of XR applications in various fields, egocentric gesture recognition is increasingly used for easy and fast human-computer communication. In the SUN project, we created a dual-stage hand gesture recognition component that uses input from the HoloLens 2 Extended Reality (XR) headset. It leverages static and dynamic neural-network models to detect "Thumbs Up", "Thumbs Down" custom static gestures, as well as "Two Moving Fingers" dynamic gesture, and translate them into corresponding commands. Initially, we collected HoloLens 2 data and trained the models for both hands. Then, we built the component, which combines the static and dynamic models. Once the static model recognizes a static "Two Fingers" gesture, the hand-joint data is buffered from the device. The buffering is terminated when the palm displacement exceeds a predefined threshold (e.g., 0.25m) in the proper direction (rightward for the right hand, leftward for the left) or after 1.2 seconds to prevent hangs. Buffered sequences are linearly resampled to a fixed length (1 sec) and then fed to the dynamic model. Users can choose to use the right or left hand for gesture input. The overall accuracy of the component is above 90%.

14.1 Introduction

Egocentric hand gesture recognition using wearable cameras—including RGB, infrared (IR), and depth sensors embedded in XR headsets—has emerged as a critical technology for natural human-computer interaction in immersive environments. These camera systems provide a rich multimodal input stream from the user's perspective, enabling gesture interpretation in real time. While RGB-only systems rely on visual appearance,

depth and IR sensors enhance robustness by providing geometric and motion cues, even under challenging lighting.

Recent methods have incorporated transformer-based architectures such as TimeSformer [Bertasius et al. 2021] to capture long-range dependencies across video frames more effectively than traditional Convolution Neural Networks (CNNs). To handle self-occlusions and hand variations, researchers have developed multimodal fusion networks that combine RGB, depth, and 3D skeleton features. In particular, Graph Convolutional Networks (GCNs) [Li et al. 2019] are used to model the spatiotemporal relationships of tracked hand joints derived from depth/IR input. Attention mechanisms [Chen et al. 2020] further improve accuracy by dynamically weighting important spatial and temporal regions. Real-time performance is achieved through lightweight models such as MobileNetV2 [Sandler et al. 2018], enabling deployment on XR headsets.

Public datasets such as EgoGesture [Zhang et al. 2018], First-Person Hand Action (FPHA) [Garcia-Hernando et al. 2018], and H2O [Kwon et al. 2021] include synchronized RGB-D(-IR) streams and annotated gestures, fostering the development of models tailored for egocentric input. Deep learning approaches such as I3D [Carreira and Zisserman 2017] and CNN-RNN hybrids are widely used to learn spatiotemporal features from these video streams.

Egocentric hand gesture recognition is increasingly used to enable natural and intuitive interaction in various domains, particularly within XR wearable systems and assistive applications. Typical use cases include hands-free control of virtual elements, such as selecting, manipulating, or navigating digital content within immersive environments. In remote assistance and training scenarios, gestures captured from a first-person view enhance task demonstration and user guidance. This technology is also valuable in industrial and clinical settings, where users can interact with digital interfaces while keeping their hands occupied. Furthermore, egocentric gesture recognition supports sign language recognition, rehabilitation monitoring, and context-aware user interfaces, enabling systems to adapt based on the user's real-time actions and environment.

In the context of the SUN project, the hand gesture recognition component extends interaction capabilities by interpreting hand gestures on the HoloLens 2 XR headset, as described in the following sections. The methods have been implemented and validated in the SUN project Case Study "Extended Reality for Safety and Social Interaction at Work" (Chapter 21).

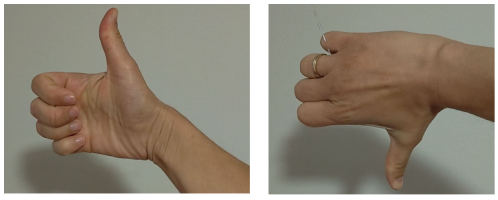


Figure 14.1: The two static gestures (a) "Thumbs Up" for "Accept" and (b) "Thumbs Down" for "Reject" are shown in the figure.

Figure 14.2: Key frames of the "Moving Two Fingers" dynamic gesture for the "Next" command.

14.2 Methodology and Results

The purpose of SUN's hand gesture recognition component is to recognize selected hand gestures of people who use the HoloLens 2 XR headset to ease human-computer interaction during the Pilot "*Extended Reality for Safety and Social Interaction at Work*". The HoloLens 2 tracks hand keypoints using a dedicated Time-of-Flight (ToF) depth sensor combined with IR cameras, processed onboard to deliver detailed, two-handed articulated hand models in real time. It can detect a set of default gestures that users perform to interact with the device (e.g., button clicks, selecting holographic items, etc.). However, the SUN's component detects a set of custom gestures, two static and one dynamic, that are not included in the default set of HoloLens 2 to help users perform some tasks more easily.

Within the framework of the SUN project, we selected the following gestures and their corresponding commands:

- "**Thumbs Up**" for "**Accept**" (see Figure 14.1a)
- "**Thumbs Down**" for "**Reject**" (see Figure 14.1b)
- "**Moving Two Fingers**" for "**Next**" (see Figure 14.2)

For the development of the SUN hand gesture recognition component, we followed the steps described below. Initially, we collected HoloLens 2 hand joints data, which

we labeled using a Unity application we created, in order to train our hand gesture recognition models for both hands (right and left). Next, we trained two deep learning neural networks for each hand: one for the static hand gestures and another for the dynamic gestures. The static model is designed to recognize the gestures "Thumbs Up," "Thumbs Down," and the static "Two Fingers" pose, while the dynamic model supports the "Moving Two Fingers" and "Random" dynamic gestures.

We subsequently constructed the SUN hand gesture component by merging the two models into a unified system. When the component is initiated, the static model is activated first. Once the static "Two Fingers" gesture is detected, the palm's initial position is recorded, and the buffering of per-frame features begins at a rate of 60Hz. To determine the appropriate moment to stop buffering hand joint data, we evaluate the palm displacement at each frame, calculated as Δ = *current position - start position*. If the magnitude of Δ exceeds a predefined threshold (e.g., 0.25m) and the movement direction matches the expected one (rightward for the right hand, leftward for the left hand), buffering is stopped immediately. If the user never crosses the displacement threshold (e.g., due to hand drifting), buffering is automatically canceled after a predefined maximum duration (1.2 seconds) to prevent the system from hanging. At that point, we have a sequence of N buffered frames, where N may be slightly shorter or longer than the maximum allowed sequence length. To standardize input to the model, we apply linear interpolation across each feature dimension to resample the sequence to a fixed length that corresponds to a duration of 1 second. As a result, the dynamic model always receives a consistent 59-frame representation. Moreover, we implemented a simple temporal smoothing mechanism so that the component only outputs "Accept" or "Reject", if the corresponding static-gesture predictions persist for several consecutive frames, rather than flickering on every single detection. This enhances the reliability of the component. In Figure 14.3, a simplified diagram of the component is shown. The following sections outline the processes of data collection, training, and validation, along with the resulting experimental findings.

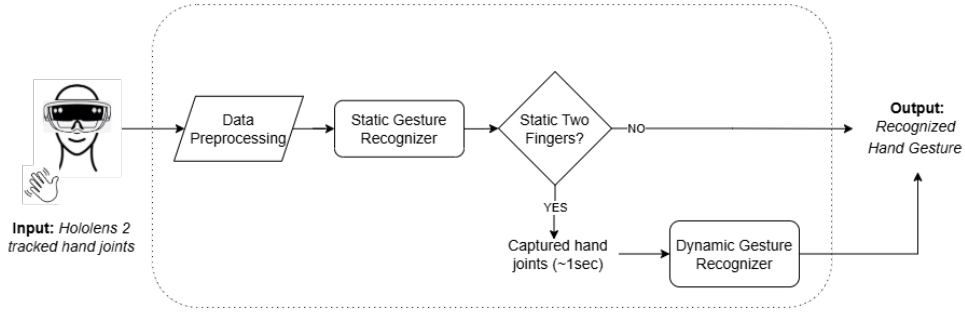


Figure 14.3: Simplified diagram of the hand gesture recognition component.

14.2.1 Data Collection and Preparation

Data Collection HoloLens 2 detects 26 hand joints per hand using its Time-of-Flight (ToF) depth sensor in combination with infrared (IR) cameras. For each joint, it provides both position and rotation information in 3D space. In our development setup, we use the Unity game engine, where the hand joint data can be accessed via Microsoft's Mixed Reality Toolkit (MRTK)[1]. To collect training data, we created a custom Unity application that captures and labels the hand joints tracked by HoloLens 2. For the static hand gesture recognition model, we collected data for four gesture categories: *"Thumbs Up"*, *"Thumbs Down"*, *"Two Fingers"*, and *"No Gesture"* (representing any other hand posture). For the dynamic gesture recognition model, we recorded one-second sequences for two gesture categories: *"Dynamic Two Fingers"* and *"Random"*. Data collection was carried out for both hands (right and left).

Concerning static data collection, the number of instances per class for each hand is presented in Table 14.1.

Table 14.1: Number of static data instances collected for each hand per gesture.

Gesture	Right	Left
No Gesture	5784	5764
Thumbs Up	5887	5700
Thumbs Down	5806	5691
Static Two Fingers	5774	5795

The number of instances per dynamic gesture for each hand dataset is shown in Table 14.2. Each instance corresponds to a dynamic gesture of one second duration.

Table 14.2: Number of dynamic data instances collected for each hand per gesture.

Gesture	Right	Left
Random	228	321
Moving Two Fingers	182	198

Data Preparation In our models, we do not use the absolute 3D positions of the 26 hand joints, as these can vary significantly depending on the hand's location in space.

[1] https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05

For example, performing the same gesture at chest level versus eye level results in substantially different XYZ coordinates for each joint, despite the hand shape being identical. This variability can hinder effective model training. To address this, we calculate the distance of each joint from the palm root, eliminating the influence of global hand position and allowing the model to focus solely on the hand's shape. Additionally, we use the raw quaternion values (x, y, z, w) for all 26 joints as provided by HoloLens 2. As a result, each input feature vector per frame comprises 25 distance values and 26 quaternions, forming a 129-dimensional vector.

14.2.2 Models' Training and Evaluation Results

In this section, we present the evaluation metrics used for assessing the SUN hand gesture recognition component, along with the training procedure of the models and the corresponding evaluation results.

Evaluation Metrics

- *Accuracy* is the proportion of all correct predictions (both true positives and true negatives) out of the total number of cases;
- *Precision* is the proportion of instances labeled positive by the model that are truly positive, reflecting how accurate its positive predictions are;
- *Recall* is the proportion of true positive cases that the model successfully identified, indicating how thoroughly it captures all actual positives;
- *F1-Score* is the harmonic mean of precision and recall, balancing the two into a single metric (high only when both precision and recall are high)

Static Model Training and Results

The static hand gesture recognition model was trained on a balanced dataset of four hand gesture classes ("NoGesture", "ThumbsUp", "ThumbsDown", and "TwoFingers"), each represented by a sequence of 129-dimensional feature vectors. These features combine 25 palm-root distances with 104 joint-orientation values (quaternion values of the 26 hand joints), all extracted from raw HoloLens 2 hand-tracking output. To ensure uniform class representation, we first loaded the training data, determined the smallest class size, and then truncated each class's data to that minimum. The resulting matrix was shuffled and split into input (x) and one-hot labels (y) before fitting a sequential Keras model with four dense layers ($129 \rightarrow 116 \rightarrow 64 \rightarrow 4$ neurons), each followed by dropout (20%), and a final softmax output.

The static model was trained for 10 epochs using the Adam optimizer and categorical cross-entropy loss, with 20% of the data held out for validation during training.

Tables 14.3 and 14.4 present the results of the static hand gesture models for the right and left hand, respectively, in terms of precision, recall, and F1-score per class on test data. The column "Number of Gestures" presents the number of static gestures used for testing per class. The overall accuracy of the trained static model for the right hand is 93.04% and for the left hand 97.37%.

Table 14.3: Per-class precision, recall, F1-score, per-class and overall accuracy (93.04%) of the static hand gesture model on test data for the right hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
NoGesture	1153	92.98%	78.06%	84.87%	78.06%
ThumbsUp	1156	91.19%	95.85%	93.46%	95.85 %
ThumbsDown	1142	99.91%	100.00%	99.96 %	100.00 %
TwoFingers	1136	88.50%	98.24%	93.12%	98.24 %
Overall	4587	—	—	—	**93.04%**

Table 14.4: Per-class precision, recall, F1-score, per-class and overall accuracy (97.37%) of the static hand gesture model on test data for the left hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
NoGesture	1165	93.28%	96.48%	94.85 %	96.48 %
ThumbsUp	1165	99.15%	100%	99.57 %	100.00 %
ThumbsDown	1170	99.19%	94.79%	96.94 %	94.79 %
TwoFingers	1129	98.05%	98.23%	98.14 %	98.23 %
Overall	4629	—	—	—	**97.37%**

Dynamic Model Training and Results:

The dynamic gesture recognition model supports two classes ("TwoFingers" and "Random"). It feeds each 129-dimension frame through two stacked TimeDistributed Conv1D blocks (64 filters, kernel size 3, each with BatchNorm + ReLU), then globally pools across the feature axis to get a (T×64) tensor. This is flattened into a single vector per sequence, passed through two small Dense + Dropout layers (128→64), and finally a 2-way softmax classifier.

Training was carried out for 18 epochs using the Adam optimizer and categorical cross-entropy loss, with 20% of the data held for validation during training.

Tables 14.5 and 14.6 present the results of the dynamic hand gesture models for the right and left hand, respectively, in terms of precision, recall and F1-score per class on test data. The column "Number of Gestures" presents the number of dynamic gestures of one-second duration used for testing per class. The overall accuracy of the dynamic models is 91.84% for the right hand and 90.00% for the left hand.

Table 14.5: Per-class precision, recall, F1-score, per-class and overall accuracy (91.84%) of the dynamic hand gesture model on test data for the right hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
Random	61	100.00%	86.89%	92.98%	91.84%
Two Fingers	37	82.22%	100.00%	90.24%	91.84%
Overall	98	—	—	—	**91.84%**

Table 14.6: Per-class precision, recall, F1-score, per-class and overall accuracy (90.00%) of the dynamic hand gesture model on test data for the left hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
Random	110	96.12%	90.00%	92.96%	90.00%
Two Fingers	40	76.60%	90.00%	82.76%	90.00%
Overall	150	—	—	—	**90.00%**

For all models, throughout training, we monitored both the training and validation loss curves to avoid overfitting; in our runs, losses decreased steadily, confirming that the networks were learning robust gesture representations without diverging. The best-performing weights for each network were exported both as an HDF5 model file and converted to ONNX for later inference in Unity.

Evaluation results of the full component: For the evaluation of the whole component, which combines the two hand gesture recognition models as shown in Figure 14.3, we used the labeled data of the three selected gestures ("Thumbs Up", "Thumbs Down" and "Moving Two Fingers"), as well as a "Random" category for any other gesture. The results are presented in Table 14.7 for the right hand and in Table 14.8 for the left hand.

Table 14.7: Per-class precision, recall, F1-score, per-class and overall accuracy (97.55%) of the hand gesture component on test data for the right hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
Random	1130	97.40 %	92.74%	95.01%	92.74%
Accept	1107	97.92%	97.92%	97.92%	97.92%
Reject	1132	98.95%	99.56%	99.25%	99.55%
Next	37	44.05%	100.00%	61.16%	100.00%
Overall	3406	—	—	—	**97.55%**

Table 14.8: Per-class precision, recall, F1-score, per-class and overall accuracy (94.55%) of the hand gesture component on test data for the left hand.

Class	Number of Gestures	Precision	Recall	F1-score	Accuracy
Random	1178	95.85%	92.11%	93.94%	92.10%
Accept	1113	99.08%	97.04%	98.05%	97.03%
Reject	1104	95.30%	99.09%	97.16%	99.09%
Next	40	55.38%	90.00%	68.57%	90.00%
Overall	3435	—	—	—	**94.55%**

14.3 Conclusions

As part of the SUN project, we developed a robust egocentric hand gesture recognition component leveraging hand joint data from the HoloLens 2 to facilitate intuitive interaction within immersive environments. The system integrates two deep learning models, one for static gesture recognition and another for dynamic gestures—designed specifically to meet the project's requirements. For a more robust component, we implemented a buffering mechanism based on displacement thresholds and time limits, followed by linear interpolation to normalize input sequences to a fixed length (1 second). Additionally, we incorporated a temporal smoothing strategy to stabilize output predictions of the static model, reducing flickering and improving the overall reliability of the system. Future work will expand the gestures' set to cover more complex hand and wrist motions and improve model accuracy through augmented datasets and advanced architectures like transformers. Another plan is to optimize on-device inference and reduce latency for smoother, lower-power operation on next-generation XR headsets.

References

Bertasius, Gedas, Heng Wang, and Lorenzo Torresani (2021). "Is Space-Time Attention All You Need for Video Understanding?" In: *Proceedings of the International Conference on Machine Learning (ICML)*. Vol. 139. PMLR, pp. 813–824.

Carreira, Joao and Andrew Zisserman (2017). "Quo vadis, action recognition? A new model and the Kinetics dataset". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6299–6308.

Chen, Liangjian, Jian Zhang, and Dacheng Tao (2020). "Gesture Recognition with Spatiotemporal Attention on 3D Skeletons". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 11068–11077.

Garcia-Hernando, Guillermo, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim (2018). "First-person hand action benchmark with RGB-D videos and 3D hand pose annotations". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 409–419.

Kwon, Taein, Bugra Tekin, Jan Stühmer, Federica Bogo, and Marc Pollefeys (2021). "H2o: Two hands manipulating objects for first person interaction recognition". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10138–10148.

Li, Maosen, Siheng Chen, Ya Zhang Zhao, Ya Wang, and Qi Tian (2019). "Actional-structural graph convolutional networks for skeleton-based action recognition". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3595–3603.

Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen (2018). "MobileNetV2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520.

Zhang, Jian, Dacheng Tao, Yong Xu, and Nicu Sebe (2018). "EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition". In: *IEEE Transactions on Multimedia* 20.5, pp. 1038–1050.

15. Multimodal Emotion Recognition

Vasileios-Rafail Xefteris[1], *Spyridon Symeonidis*[1], *Sotiris Diplaris*[1], *and Stefanos Vrochidis*[1]

[1] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece

Abstract. This chapter describes the multimodal emotion recognition developed in the context of the SUN project. The multimodal emotion recognition is based on the development and fusion of two different unimodal components, the sensor-based and the visual-based emotion recognition models. The sensor-based component is based on physiological data from an EmotiBit wearable device, using a typical machine learning process. The computer vision model was developed by adopting the MobileNetV2 architecture. In order to simulate the use of a VR mask, the publicly available AffectNet dataset was modified, masking the subjects' eyes. The multimodal fusion is based on a decision-level approach, fusing the outputs of the unimodal components. For the training of the sensor-based component, a data collection was performed to acquire annotated data using an EmotiBit wearable device. During the same data collection protocol, facial expression videos of the users were also captured, obtaining synchronized sensor and visual data that were used for the training of the multimodal fusion component. The multimodal emotion recognition component and the sensor-based emotion recognition component are used in the SUN projet Case Study "Extended Reality for Rehabilitation" and "Extended Reality for People with Serious Mobility and Verbal Communication Diseases" respectively.

15.1 Introduction

Emotion recognition is the process of predicting, analyzing, and processing human emotions by utilizing different types of multimodal data, including facial expressions and physiological responses. Emotion recognition from facial features in Extended Reality (XR) environments presents unique challenges due to the partial occlusion caused by head-mounted displays, which obscure most of the upper face, including the eyes and forehead regions critical for affective expression. Traditional computer vision models trained on fully visible faces tend to underperform in this context. Recent research has focused on adapting or retraining deep learning architectures (e.g., Convolutional Neural Networks (CNNs), MobileNet, or ResNet variants) using occluded or synthetic datasets to improve robustness under limited visual input [Li et al. 2018]. Techniques such as data augmentation with artificial occlusion, facial landmark estimation under occlusion, and multimodal fusion (e.g., combining facial features with body posture, speech or physiological signals) have shown promising results [Casas-Ortiz et al. 2024]. Lightweight models (e.g., MobileNetV2 [Sandler et al. 2018]) combined with real-time face tracking (e.g., via MediaPipe [1] or OpenFace [2]) are increasingly used to maintain performance in embedded XR systems. However, accurately recognizing emotions from the lower face alone—primarily mouth and jaw movements—remains an open research problem, especially for subtle expressions and valence/arousal estimation.

The nature of emotions as physiological reactions to different stimuli has led to the extensive use of physiological sensors in the field of emotion recognition. Among the most common physiological signals used for emotion recognition are electrodermal activity (EDA), photoplethysmography (PPG), electrocardiograph (ECG), and electroencephalography (EEG) [Egger et al. 2019]. The complex nature of emotions and the way they are depicted on physiological signals have led to the emergence of solutions that utilize a combination of multiple physiological signals, taking advantage of their complementary nature [Xefteris et al. 2022].

The complementary nature of facial expressions and physiological signals has been revealed through multimodal fusion solutions. Multimodal solutions that fuse physiological signals and facial expressions using machine learning [Moin et al. 2023] and deep learning [Jung, Sejnowski, et al. 2019] have shown the superiority of fusing both modalities rather than utilizing only unimodal data. Such solutions can improve unimodal components performance and also deal with the problems from each component, like the occlusion problems of computer vision solutions.

[1] https://mediapipe.readthedocs.io/en/latest/solutions/face_detection.html
[2] https://github.com/TadasBaltrusaitis/OpenFace

In the context of the SUN project, the multimodal emotion recognition is a component implemented in the Case Study "Extended Reality for Rehabilitation" (Chapter 20), while the sensor-based emotion recognition is a component used in the Case Study "Extended Reality for People with Serious Mobility and Verbal Communication Diseases" (Chapter 22). The components are responsible for capturing and estimating the participants' emotional states while they perform different actions in the context of each Case Study.

15.2 Methodology and Results

The current section provides a description of the approach, experiments, and findings regarding the multimodal fusion process of emotion recognition of the SUN project.

15.2.1 Emotion Identification Based on Facial expressions

SUN's visual-based emotion recognition component estimates users' emotional states in real time by analyzing visible facial features captured from an external RGB camera. This component is used in the SUN Case Study "Extended Reality for Rehabilitation" (Chapter 20). Its aim is to provide additional emotional information to the multimodal system, offering insight into how users feel while performing physical exercises. The presence of the XR headset introduces a significant challenge, as it partially occludes the upper part of the face, limiting the available visual information. To address this, the component is built so that it exploits only the lower part of the face to extract the emotion information. Its output is subsequently sent to the multimodal emotion recognition module, as illustrated in Figure 15.1. The ultimate goal is to support the dynamic adaptation of the XR experience to better accommodate each user's emotional and physical state.

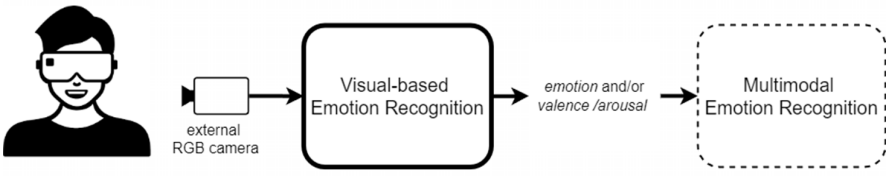


Figure 15.1: Visual-based emotion recognition diagram.

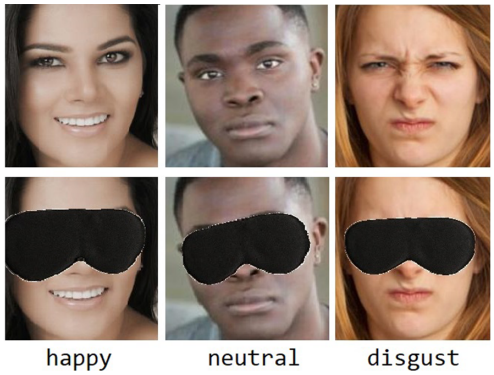


Figure 15.2: Example images from both the original and masked AffectNet dataset.

Dataset: More specifically, in order to train the emotion identification models used in this component, we created a masked version of the AffectNet dataset [Mollahosseini et al. 2017] to simulate XR occlusion. AffectNet contains about 1M facial images each one labeled with one of 8 emotional states (*neutral, happy, angry, sad, fear, surprise, disgust, contempt*) along with the intensity of valence and arousal (values between -1 and 1). With the help of the MaskTheFace[3] GitHub project, which was adapted to our needs, we masked the upper face of the original AffectNet images. Figure 15.2 presents some example masked images.

Training of the Models

We trained the MobileNetV2 architecture [Sandler et al. 2018] on the masked AffectNet dataset. We adapted the output layers of the MobileNetV2 and tested different training parameters, resulting in the following three emotion recognition models that were trained using our masked AffectNet dataset:

- "*MobileNetV2 - 8 Emotions*", which recognizes 8 emotion categories;
- "*MobileNetV2 - VA*" that outputs the valence and arousal continuous values;
- "*MobileNetV2 - 8 emotions + VA*", which recognizes both the 8 specific emotions and the valence and arousal values.

All the above models have been pre-trained on the ImageNet dataset [Deng et al. 2009] and then trained using our masked AffectNet dataset. The derived models

[3] https://github.com/aqeelanwar/MaskTheFace

revealed a clear bias toward majority classes like “happy” and “neutral.” This imbalance likely stems from the underlying class distribution in AffectNet.

To address the class imbalance in AffectNet, we re-trained the MobileNetV2 model using class-aware oversampling. Specifically, we computed per-sample probabilities inversely proportional to each emotion’s frequency and used these as weights in the data generator during training. Instead of uniformly shuffling, the generator resamples with replacement based on these weights at each epoch, ensuring that minority emotions like “disgust” and “contempt” appear as frequently as dominant ones like “happy” and “neutral”. Additionally, we applied data augmentation to oversampled images using mild geometric and photometric transformations—such as flips, affine shifts, and brightness or contrast adjustments—to enhance variability without distorting facial expressions. These measures helped balance each mini-batch and encouraged the model to learn robust features across all emotion categories, reducing overfitting to majority classes.

Evaluation Results

Table 15.1 presents evaluation results for our three MobileNetV2 variants, using the validation set of the masked AffectNet dataset (3,999 images), since AffectNet’s test set is not publicly available. For emotion classification, accuracy was used as the evaluation metric, while valence and arousal predictions were evaluated using Root Mean Square Error (RMSE). Accuracy indicates the percentage of correct emotion predictions (0–100%), while RMSE measures the average Euclidean distance between predicted and true values, with lower values indicating better performance.

SUN’s MobileNetV2 models were trained on: (i) the original AffectNet dataset, (ii) a version with simulated occlusion (masked AffectNet), and (iii) a class-balanced version of masked AffectNet. Each training set consisted of 287,651 images. All of the above models were evaluated on the same images, those of the masked AffectNet validation set. For reference, the original AffectNet paper reported a baseline accuracy of 58% and average RMSE of 0.39 for valence and arousal—achieved using unmasked validation images. Our ”MobileNetV2 - 8 Emotions” model trained on the masked AffectNet dataset reached an accuracy of 51.8% on the masked AffectNet validation set, which is the best value we got. For the estimation of valence and arousal, the ”8 Emotions + VA” model is the best, giving an average RMSE of 0.35 (0.367 for valence and 0.342 for arousal). In addition, Table 15.1 presents the performance of the three models on the original (unmasked) AffectNet validation set, highlighting how a model’s performance improves as it learns to rely more effectively on the lower facial region to extract the affective information.

Table 15.1: Evaluation of the visual-based emotion recognition MobileNetV2-based models on the validation set of the original, masked and balanced masked AffectNet in terms of Accuracy and RMSE.

Model	Dataset	Acc. (%)	Valence RMSE	Arousal RMSE
8 Emotions	Original	33.7	–	–
	Masked	**51.8**	–	–
	Masked (balanced)	49.9	–	–
VA	Original	–	0.443	0.510
	Masked	–	0.458	0.380
	Masked (balanced)	–	0.404	0.350
8 Emotions + VA	Original	27.9	0.435	0.369
	Masked	41.9	0.438	0.368
	Masked (balanced)	50.5	**0.367**	**0.342**

Implementation of the Component:

To implement the component, we developed a Python script that runs the selected emotion recognition model in real time using input from an external RGB camera. MediaPipe's face detector[4] isolates the face region, which is then fed into the emotion recognition model. Depending on the selected model variant, the component outputs the predicted emotion label with its confidence score, the valence/arousal values, or both. This output is passed directly to the multimodal fusion module through the Upper Limbs pipeline in the SUN Case Study "Extended Reality for Rehabilitation" (Chapter 20).

15.2.2 Sensor-Based Emotion Recognition

This application includes both the analysis of wearable data for emotion recognition and the development of algorithms for the multimodal fusion of wearable and visual data for the same cause. The multimodal emotion recognition is based on visual data from facial videos and physiological signals from an EmotiBit wearable sensor.

Methods

The sensor-based emotion recognition module analyzes physiological signals acquired from the deployed EmotiBit wearable sensor using a typical feature extraction process.

[4] https://mediapipe.readthedocs.io/en/latest/solutions/face_detection.html

From the EmotiBit sensor, three different physiological signals are collected in order to be analyzed; those being photoplethysmograph (PPG), electrodermal activity (EDA) and temperature (TEMP) data. The extracted features are as shown in Table 15.2.

Table 15.2: Extracted Features from PPG, EDA, and TEMP Signals

Feature	Description
	PPG
Average HR (BPM)	The mean heart rate calculated from detected peaks in beats per minute.
SDNN (ms)	Standard deviation of NN intervals (the time between successive heartbeats).
RMSSD (ms)	Root mean square of successive differences between adjacent NN intervals.
pNN50 (%)	Percentage of successive NN intervals that differ by more than 50 ms.
LF Power	Power in the low-frequency band of the HRV spectrum.
HF Power	Power in the high-frequency band of the HRV spectrum.
LF/HF Ratio	Ratio of low-frequency to high-frequency power.
Sample Entropy	A measure of complexity and irregularity in the time series of inter-beat intervals.
Mean Pulse Amplitude	Average amplitude of the detected PPG peaks.
Std Pulse Amplitude	Standard deviation of the pulse amplitude.
Respiration Rate (BPM)	Estimated rate of breathing derived from the PPG signal.
	EDA
Mean	The average value of the EDA signal.
Standard Deviation (std)	The standard deviation value of the EDA signal.
Variance (var)	The variance value of the EDA signal.
Minimum (min)	The lowest value in the EDA signal.
Maximum (max)	The highest value in the EDA signal.
Skewness	A measure of the asymmetry of the EDA signal distribution.
Kurtosis	A measure of the "tailedness" of the EDA signal distribution.
Num Peaks	The number of significant peaks detected in the EDA signal.
Mean Peak Amplitude	The average amplitude of detected SCR peaks.

Feature	Description
Peak Variance	The variance of the peak amplitudes.
PSD Mean	The mean value of the power spectral density.
PSD Max	The maximum value of the power spectral density.
PSD Sum	The total power across all frequencies in the power spectral density.
Entropy	Shannon entropy of the EDA signal.
	TEMP
Mean	The average temperature value over the signal duration.
Standard Deviation (std)	The standard deviation of temperature over the signal duration.
Variance (var)	The variance of temperature over the signal duration.
Minimum (min)	The lowest temperature value observed in the signal.
Maximum (max)	The highest temperature value recorded in the signal.
Skewness	A measure of the asymmetry of the temperature distribution.
Kurtosis	A measure of the "tailedness" of the temperature distribution.
Num Peaks	The number of significant peaks detected in the temperature signal.
Mean Peak Amplitude	The average amplitude of detected temperature peaks.
Peak Variance	The variance of the peak amplitudes.
PSD Mean	The mean value of the power spectral density.
PSD Max	The maximum value of the power spectral density.
PSD Sum	The total power across all frequencies in the power spectral density.
Entropy	Shannon entropy of the temperature signal.

All features are extracted using a sliding window technique with 30 seconds length and 50% overlap. For the fusion of the different physiological signals the extracted features are concatenated and fed to a trained machine learning model in order to predict valence and arousal scores.

Finally, the multimodal fusion component for emotion recognition receives as input the predictions of the unimodal emotion recognition components, synchronizes them and fuses them to a final unified outcome.

Dataset

In order to train the multimodal fusion component it is important to have synchronized annotated data of all deployed modalities. For this cause, a data collection experiment

was designed in order to collect synchronized data from the EmotiBit sensor and facial video data. During the experiment, participants were asked to watch a series of 16 small video clips from well-known movies and evaluate them in regard to the elicited emotion. For this evaluation, the valence-arousal 2D emotional representation was selected. Users' scores range from -3 to 3 for both valence and arousal metrics. During the whole experiment, participants were wearing an EmotiBit wearable sensor on their wrist to monitor their physiological signals while also a camera was used to capture their facial videos.

Results

For the sensor-based emotion recognition component, four different machine learning models, namely Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Decision Tree and Random Forest, were tested regarding their accuracy in the prediction of valence and arousal scores. The results are presented in Table 15.3, revealing the superiority of Random Forest over the rest of the models tested.

Table 15.3: Accuracy results of different machine learning algorithms on valence and arousal scores based on the collected dataset.

Model	Valence score	Arousal score
SVM	24.52	31.64
kNN	39.54	41.27
Decision tree	81.41	78.86
Random Forest	91.71	90.84

15.2.3 Multimodal Fusion

Since both sensor-based and computer-vision-based models are deployed for emotion recognition, there is a need to develop a multimodal fusion component that effectively combines the two modalities into a unified outcome. The methods adopted were based on a decision-level architecture, meaning that the results from the two different unimodal components are fused in a late fusion manner. The two different methods are averaging and accuracy-weighted fusion. The former one is a simple averaging of the probabilities for each class produced by the unimodal components, while the latter one is a weighted averaging method with the accuracy scores for each element being the weight of this component. The results of the two different fusion methods are presented in Table 15.4, revealing the superiority of the accuracy-weighted fusion method over both the unimodal components and the averaging method.

Table 15.4: Accuracy results of different multimodal fusion techniques on valence and arousal scores based on the collected dataset.

Model	Valence score	Arousal score
Sensor based	91.71	90.84
Averaging fusion	85.74	73.57
Accuracy-weighted fusion	94.44	95.26

15.3 Conclusions

This chapter describes the process of multimodal emotion recognition developed in the context of the SUN project. The component is based on two different unimodal components, the sensor-based and the visual-based components, followed by a multimodal fusion component.

The visual-based emotion recognition component enables real-time estimation of users' emotional states based on facial features captured by an external RGB camera. To address the occlusion introduced by XR headsets, the model was trained with artificial occlusion augmentation and optimized using a lightweight MobileNetV2 architecture. This approach increases the emotion classification and valence/arousal estimation performance in the case of XR headset face occlusion compared to the models trained on non-occluded images. The component is integrated into the XR pipeline to provide affective feedback, supporting the dynamic adaptation of the experience to individual users. Future work will focus on improving robustness and incorporating temporal emotion dynamics. The sensor-based component is based on an EmotiBit wearable device, which was also used for the collection of a multimodal dataset for emotion recognition. This dataset was used to train the sensor-based emotion recognition component. A typical machine learning approach was followed, with future work focusing on more advanced deep-learning approaches.

The two different modalities were fused using a standard decision-level approach, using an accuracy-weighted fusion as the better-performing method. Future work will focus on using more advanced techniques, including attention mechanisms in the fusion process.

References

Casas-Ortiz, Alberto, Jon Echeverria, Nerea Jimenez-Tellez, and Olga C Santos (2024). "Exploring the impact of partial occlusion on emotion classification from facial expressions: A comparative study of XR headsets and face masks". In: *IEEE Access* 12, pp. 44613–44627.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Egger, Maria, Matthias Ley, and Sten Hanke (2019). "Emotion recognition from physiological signal analysis: A review". In: *Electronic Notes in Theoretical Computer Science* 343, pp. 35–55.

Jung, Tzyy-Ping, Terrence J Sejnowski, et al. (2019). "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing". In: *IEEE Transactions on Affective Computing* 13.1, pp. 96–107.

Li, Yong, Jiabei Zeng, Shiguang Shan, and Xilin Chen (2018). "Occlusion aware facial expression recognition using CNN with attention mechanism". In: *IEEE transactions on image processing* 28.5, pp. 2439–2450.

Moin, Anam, Farhan Aadil, Zeeshan Ali, and Dongwann Kang (2023). "Emotion recognition framework using multiple modalities for an effective human–computer interaction". In: *The Journal of Supercomputing* 79.8, pp. 9320–9349.

Mollahosseini, Ali, Behzad Hasani, and Mohammad H Mahoor (2017). "Affectnet: A database for facial expression, valence, and arousal computing in the wild". In: *IEEE Transactions on Affective Computing* 10.1, pp. 18–31.

Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen (2018). "MobileNetV2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520.

Xefteris, Vasileios-Rafail, Athina Tsanousa, Nefeli Georgakopoulou, Sotiris Diplaris, Stefanos Vrochidis, and Ioannis Kompatsiaris (2022). "Graph theoretical analysis of EEG functional connectivity patterns and fusion with physiological signals for emotion recognition". In: *Sensors* 22.21, p. 8198.

SUN Platform and Cybersecurity

The SUN Platform embodies a comprehensive approach for secure, interoperable, and human-centered Extended Reality environments. At its foundation lies an integrated technological framework that integrates diverse components through advanced middleware, enabling seamless communication, scalable deployment, and robust cybersecurity aspects. Within this ecosystem, blockchain-based mechanisms ensure trustworthy digital asset management through tokenization, smart contracts, and decentralized data governance, reinforcing transparency and data sovereignty. Complementing these capabilities, an intelligent cyber threat detection system safeguards users during immersive experiences, providing real-time protection against security breaches. Together, these innovations establish a comprehensive foundation for reliable, ethical, and resilient XR applications.

16. SUN Integrated Platform

Alexandru Stan[1], *Preslav Rachev*[1], *George Ioannidis*[1], *Ferdinando Bosco*[2], *and Vincenzo Croce* [2]

[1] IN2 Digital Innovations, Germany
[2] Engineering Ingegneria Informatica S.p.a. (ENG), Italy

Abstract. The SUN Integrated Platform represents an innovative technological framework designed to enhance Extended Reality environments by integrating diverse functionalities within a secure, scalable, and user-centric architecture. At its core, the platform features OmniBridge, an advanced middleware solution facilitating efficient communication, service discovery, and robust data management across multiple XR components. OmniBridge supports MQTT (Message Queuing Telemetry Transport) and gRPC (gRPC Remote Procedure Calls) protocols, optimizing secure data exchanges and significantly enhancing system observability through detailed tracing capabilities. The platform offers flexible deployment options, accommodating local, cloud, and hybrid environments, thus catering to various professional contexts. Through a well-defined modular architecture, the SUN Integrated Platform enables seamless integration and interaction among 26 distinct technical components, providing comprehensive digital asset tokenization, real-time cyber threat detection, and data persistence. These advanced capabilities collectively improve operational efficiency, asset security, and user safety, making the platform particularly suitable for professional use across diverse disciplinary fields. The comprehensive integration of cutting-edge cybersecurity measures and blockchain technologies, alongside the ethical and user-centered approach used in the design and validation phases, further underscores the platform's commitment to human-centered, secure, and reliable XR applications.

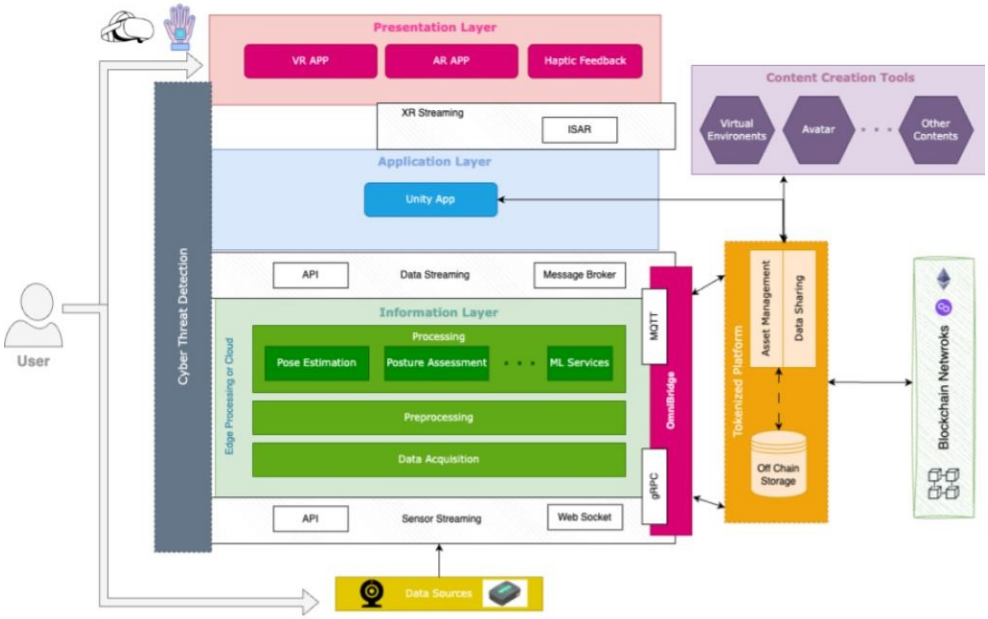


Figure 16.1: The SUN Platform architecture.

16.1 Introduction

16.1.1 Context and Motivation

At the heart of the SUN project lies a modular architecture designed to combine cutting-edge wearable devices, artificial intelligence, and immersive technologies into a seamless whole. This is the SUN XR Integrated Platform, containing all of the foreground components developed by SUN partners within the project (see Figure 16.1). You might think of it as a digital ecosystem where cameras, sensors, and headsets collaborate to make virtual experiences feel more lifelike, while robust security and content management tools ensure that every piece of digital content is handled safely and transparently. From rehabilitation clinics to factory floors, the SUN platform is already being tested in real-world situations like physiotherapy, workplace safety training, and support for people with disabilities.

The *Architectural Layers (Information, Application, and Presentation)* are the core part of the SUN Reference Architecture, in which all the XR components can be developed and integrated, respecting their own main characteristics.

The *Communication Layers (Sensor, Data, and XR streaming)* have different roles and capabilities, depending on the needs of the connected Architectural Layers. They implement and integrate different protocols, ensuring the best performance at any level of the process. In addition, it was also clear that the necessity to ensure security and privacy aspects during the data sharing, and for this reason, two core components are shared in the overall platform for ensuring data storage, access management, and cybersecurity: the *Cyberthreat Detection* and the *Tokenized Platform*.

The *Content Creation Tools* were grouped as external tools since they can work offline and outside of the deployed platform. They can be easily connected through the Tokenized Platform, ensuring data access management and data sharing in a scalable and secure way.

The SUN Integrated Platform emerges from a growing necessity to harness the capabilities of XR technologies across diverse professional applications, necessitating an infrastructure capable of securely managing complex digital environments. XR technologies, encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), provide immersive and interactive experiences but pose significant challenges in cybersecurity, asset management, and system integration.

16.1.2 Challenges in Traditional XR Systems

Traditionally, XR systems have struggled with fragmented hardware and software ecosystems. Different vendors and tools often use incompatible protocols and data formats, making it difficult to connect devices and components into a cohesive, scalable environment. This forces developers and engineers to use brittle, time-consuming, and difficult-to-maintain custom integrations. Furthermore, existing XR systems often lack cybersecurity protections designed specifically for immersive technologies. Beyond standard data breaches, attacks against XR platforms have the potential to alter user perception or even result in physical disorientation. Such flaws make it very hard to use XR in places where safety is important, like in industrial training or patient rehabilitation. Lastly, performance and latency issues are common in XR systems, particularly when real-time responsiveness is required. Fragmented architectures can introduce lag or processing delays that compromise the immersive experience or interfere with synchronized collaboration across users. These gaps in interoperability, security, and performance have slowed the uptake of XR technologies in professional environments.

The SUN Integrated Platform was developed in direct response to these limitations. It provides a secure, standards-based foundation for integrating diverse XR components while embedding real-time threat detection, latency-aware data flows, and robust cybersecurity into the system architecture.

16.1.3 Objectives of the SUN Integrated Platform

At the heart of this integrated approach is OmniBridge, a middleware solution that acts as the central backbone for communication and coordination among various platform components. OmniBridge enhances interoperability, security, and operational transparency, supporting complex data exchanges and providing advanced monitoring and debugging tools.

The platform also incorporates blockchain technology for digital asset tokenization, providing secure and transparent management of XR assets via NFTs (Non-Fungible Tokens)[1]. Furthermore, the integration of an advanced Cyber Threat Detection module addresses the inherent vulnerabilities within XR environments, leveraging real-time machine learning algorithms to ensure user safety and system integrity.

The SUN Integrated Platform thus offers professionals from multidisciplinary fields a comprehensive, secure, and adaptable infrastructure, enabling them to leverage the full potential of XR technologies. By addressing the fragmented nature of current XR solutions, SUN fosters greater efficiency, improved user experiences, and heightened security, laying a solid foundation for future advancements in XR applications.

16.2 Architecture Overview

The architecture of the SUN platform is built around a pragmatic need: to allow distributed XR components, such as sensors, analysis tools, and immersive applications, to collaborate in real time without manual integration. To achieve this, the system relies on a single middleware layer, OmniBridge, which handles registration, session management, and secure data exchange between components. Communication is organized around gRPC[2] APIs and MQTT topics, with each session isolated and traceable. In addition to this foundation, two key modules extend the platform's capabilities: a tokenization system that manages digital asset ownership and off-chain data storage, and a Cyber Threat Detection module that protects XR users from immersive attacks in real time. Both modules are introduced in this chapter in connection with the overall architecture, while their design and implementation are examined in depth in the following dedicated chapters.

[1] https://en.wikipedia.org/wiki/Non-fungible_token
[2] https://grpc.io

16.3 Omnibridge: Integration Middleware

At the core of this ecosystem is OmniBridge, the middleware glue that binds every component together. Imagine a central control room that knows which devices are available, which SUN applications are running, and how data should flow between modules — this is precisely what OmniBridge provides. Developed as a central hub for service discovery and secure communication, OmniBridge registers each module, assigns it a unique identifier, and issues cryptographic credentials so that sensors, AI algorithms, and streaming services can exchange data without risk of eavesdropping or interference. On top of this, it provides a high-performance data exchange and monitoring. OmniBridge itself is built as three collaborating services:

- A *Core API* for component registration, authentication, session management, and a dynamic catalog;
- An *MQTT message broker* optimized for high-throughput, real-time data streaming with custom security plugins;
- A *Monitoring service* that provides distributed tracing and observability.

By combining synchronous control calls (via high-performance gRPC interfaces) with asynchronous data streams (via lightweight MQTT topics), OmniBridge achieves both reliability and speed, allowing, for example, a motion sensor to broadcast dozens of readings per second while another component listens and responds without delay.

16.3.1 Architecture and Communication Protocols

OmniBridge uses a three-service architecture that changes how XR components communicate. The core OmniBridge service exposes gRPC API endpoints for component registration, authentication, session management, and catalog access. Unlike traditional REST APIs, gRPC uses Protocol Buffers to serialize data into a compact binary format, resulting in faster data transmission and lower network usage compared to text-based JSON. The advantages of gRPC over REST are significant. Protocol Buffer definitions ensure strongly-typed service contracts and message structures, reducing errors and improving reliability. The system supports automatic client and server code generation in multiple languages, ensuring consistency across development teams. Additionally, gRPC natively supports bi-directional streaming for real-time communication, making it suitable for XR applications requiring continuous data flow.

For real-time data exchange, OmniBridge integrates a customized Eclipse Mosquitto MQTT message broker. MQTT enables scalable asynchronous communication through

a lightweight, publish/subscribe messaging protocol optimized for unreliable networks and low-bandwidth devices. This dual-protocol approach ensures reliable service management through gRPC and high-throughput, real-time data exchange through MQTT.

gRPC and Protocol Buffers over REST

The choice of gRPC and Protocol Buffers represents a fundamental shift from traditional REST-based communication patterns that still dominate service-to-service communication on the Internet. Understanding this technology stack helps explain why OmniBridge achieves its performance and reliability characteristics.

Protocol Buffers: Schema-First Communication: Protocol Buffers, developed by Google, provide a language-neutral, platform-neutral mechanism for serializing structured data. Unlike JSON-based REST APIs, where message structure is often implicit or loosely defined, Protocol Buffers require explicit schema definitions that specify exact message formats, field types, and service contracts. This schema-first approach brings several advantages to distributed XR systems. Every component knows exactly what data structures to expect, eliminating the runtime errors common with loosely-typed JSON payloads. When a motion sensor sends posture data or a haptic device receives force feedback commands, both sender and receiver use identical, compiler-verified data structures.

The shared Protocol Buffers specification lives in OmniBridge's integrated repository, ensuring all partners access the same version. Changes go through pull request reviews, maintaining specification consistency while enabling collaborative evolution. This collaborative approach prevents the API drift that often plagues distributed systems as they scale.

Binary Serialization Performance: The performance difference between Protocol Buffers and JSON becomes critical in XR applications. While REST APIs serialize data into human-readable text, Protocol Buffers create compact binary representations. For multiple components that broadcast dozens of sensor readings per second, this efficiency translates directly into reduced network usage and faster transmission times. Consider a typical XR scenario where multiple components need to report data quickly. A heart rate monitor might broadcast at 60Hz, while motion sensors stream positional updates at even higher frequencies. While they don't broadcast every single message using OmniBridge's gRPC APIs, this might very well become the norm in the near future, which is why we have prepared a solid foundation. The cumulative bandwidth savings from binary serialization become substantial as component counts increase.

Code Generation and Multi-Language Support: Protocol Buffers support automatic code generation across programming languages—from Go and Java to Python and

JavaScript. This means XR components written in different languages can communicate seamlessly without manual interface adaptation. A Python-based AI module can exchange strongly-typed messages with a C++ rendering engine, with both sides guaranteed to use compatible data structures. The code generation eliminates human error in API implementation. Instead of manually writing serialization code that might introduce subtle bugs, developers use compiler-generated functions that are guaranteed to match the schema specification. This reliability becomes crucial when debugging complex XR workflows where multiple components interact simultaneously.

A Look to the Future - gRPC's Streaming Capabilities: Beyond simple request-response patterns, gRPC natively supports bi-directional streaming. This capability proves essential for XR applications requiring continuous data flow. A gesture recognition component can stream hand tracking data while simultaneously receiving calibration updates, all within a single connection. The streaming support also enables more sophisticated communication patterns. Components can establish persistent connections for low-latency data exchange while still using standard RPC calls for configuration and control operations. This flexibility allows OmniBridge to optimize communication patterns based on specific use cases.

Implementation with OmniBridge: Within OmniBridge, gRPC handles the "control plane" operations—component registration, session management, and catalog queries—where reliability and strong typing matter most. Meanwhile, MQTT handles the "data plane" for high-frequency sensor streams where throughput and low latency take priority. This dual-protocol approach leverages each technology's strengths while avoiding its respective limitations. The Protocol Buffers schema serves as a contract between OmniBridge and all connected components. When new functionality is added or existing APIs evolve, the schema changes provide clear documentation of what has changed and how it impacts existing integrations. This contract-driven development approach scales effectively as the platform grows and supports increasingly complex XR scenarios.

Messaging Communication Over MQTT

While gRPC handles the structured, reliable control operations in OmniBridge, MQTT serves as the high-throughput data exchange mechanism that enables real-time communication between XR components. Understanding MQTT's publish/subscribe model helps explain how OmniBridge achieves its benchmark performance of 600-800 messages per second for asynchronous communication.

The Publish/Subscribe Paradigm: MQTT operates on a fundamentally different communication model than traditional request-response protocols. Instead of components maintaining direct connections with each other, they connect to a central message

broker—in OmniBridge's case, a customized Eclipse Mosquitto instance. Components can publish messages to named topics or subscribe to topics of interest, with the broker handling message routing automatically. This decoupling proves essential for XR applications where components have varying data production and consumption patterns. One component may publish messages at the speed of hundreds or thousands per second, while other subscribe to those readings simultaneously. The publisher remains unaware of how many consumers exist or their individual processing capabilities.

Hierarchical Topic Organization: OmniBridge structures MQTT communication through a three-level topic hierarchy that balances global coordination with session isolation. Global broadcasts use the broadcast topic for system-wide notifications. Scope-specific broadcasts follow broadcast/[SCOPE_TAG] for location-bound messaging, allowing components in the same physical room to coordinate without broadcasting platform-wide. Component-specific broadcasts use broadcast/[SCOPE_TAG]/[COMPONENT_ID] for targeted notifications. For active sessions, each component receives a dedicated topic following sessions/[SESSION_ID]/[COMPONENT_ID]. This namespacing ensures secure, isolated data exchange between session participants while preventing data leakage between concurrent experiments, i.e., application runs.

Custom Topic Configuration and Experimental Bridging: Omnibridge provides flexible topic naming to address pilot feedback. Components can register with predefined, semantically meaningful topic names while maintaining isolation guarantees. Experimental MQTT bridging has been locally tested to enable distributed deployments by allowing multiple brokers to forward messages transparently. This could support hybrid scenarios where local XR sessions communicate with remote services, although bridging requires shared JWT validation across brokers and careful consideration of network latency. While not available to all integrators, it shows potential for further investigation.

Security and Authentication

Security in XR platforms presents unique challenges due to the immersive nature of the technology and the sensitive data involved. OmniBridge addresses these concerns through a comprehensive multi-layer security architecture that evolved significantly between its first and second releases.

JWT-Based Authentication: The foundation of OmniBridge's security model relies on JSON Web Tokens for stateless authentication and authorization. Each component receives a unique Component ID and Component Secret during registration, which it uses to obtain a cryptographically signed JWT token. These tokens embed component identity and permission claims that dictate exactly which MQTT topics the component can publish to or subscribe from. This approach eliminates the need for repeated calls

to centralized authentication servers, supporting scalable, low-latency access control across distributed XR components.

MQTT-Level Security Implementation: OmniBridge integrates a custom low-level plugin into the Eclipse Mosquitto broker that intercepts every incoming message and performs real-time security checks. The plugin validates JWT tokens in message headers and inspects permission claims to determine whether the sender is authorized for specific publish or subscribe operations. Despite operating at the message level, this security implementation maintains high performance with no measurable latency or throughput degradation under high load.

Tracing and Observability

The asynchronous nature of XR component communication creates unique debugging challenges that traditional logging cannot address effectively. When multiple components interact simultaneously—a motion sensor streaming data while an AI module processes gestures and a haptic device provides feedback—understanding the flow of operations across these distributed services requires more sophisticated tooling than simple log files.

OmniBridge addresses this challenge through distributed tracing, a strategy adapted from microservice architectures. Each session receives a unified tracing ID that propagates across all participating components, while individual components define unique span IDs to mark their participation boundaries. This approach creates a coherent timeline visualization where developers can inspect complex workflows, identify performance bottlenecks, and trace data flow across the entire system. The observability UI, built on the Jaeger[3] tracing framework, displays session lifecycles as interactive timelines showing operations like session management, MQTT data exchanges, and component interactions (see Figure 16.2). The system includes persistent trace storage, ensuring session traces survive system restarts and remain available for future analysis. Developers can export traces as JSON files for offline inspection and collaborative debugging, then re-import them later for detailed investigation. The system also integrates MQTT throughput statistics into the trace timeline, enabling correlation between data volume anomalies and reported issues without the overhead of tracing every individual message.

Performance Optimization

OmniBridge is optimized to support high-performance demands, balancing efficient resource utilization with robust system reliability. Its architecture facilitates efficient scala-

[3] https://www.jaegertracing.io/

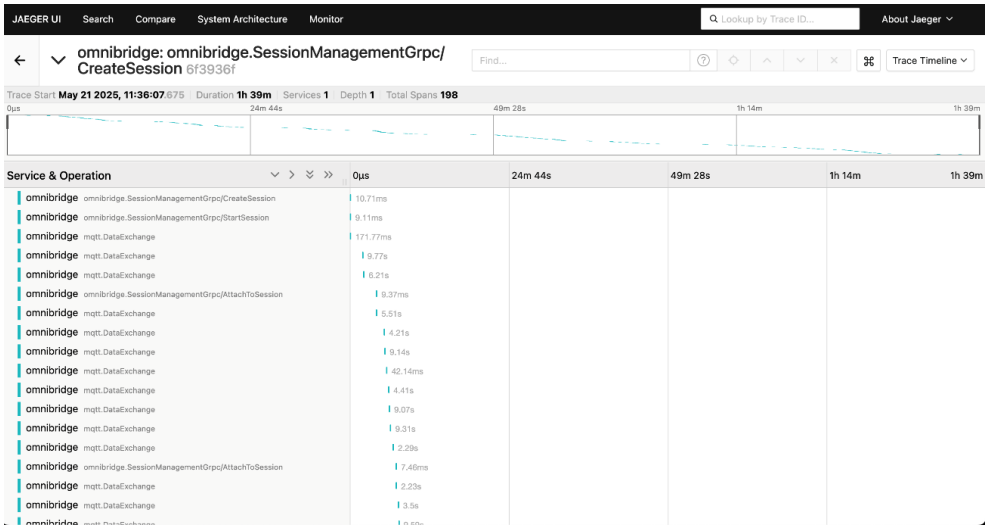


Figure 16.2: The Jaeger Observability System

bility, accommodating increased loads without compromising performance. Benchmark testing demonstrates OmniBridge's ability to maintain low latency and high throughput, essential for real-time XR applications. Through these advanced functionalities, OmniBridge middleware ensures robust, secure, and scalable interactions across all integrated components, providing a resilient foundation for complex XR ecosystems within the SUN Integrated Platform.

16.3.2 Tokenized Platform

The Tokenized Platform manages ownership and data storage within the SUN ecosystem. It runs on blockchain and standard protocols, allowing users to register, manage, and share XR data and assets. The system includes two parts: XR Asset Tokenization and Off-Chain Storage. The first sets ownership and access using smart contracts and NFTs. The second serves as a data repository for the SUN Platform. Integrated with OmniBridge, it handles storage, logging, and retrieval of platform data, contributing to the foundation of a European Dataspace for digital assets. More details on the Tokenized Platform will be provided in Chapter 17 .

Asset Tokenization

The XR Asset Tokenization module is based on Ethereum-compatible smart contracts, supporting both fungible tokens (ERC-20) and non-fungible tokens (ERC-721). Content

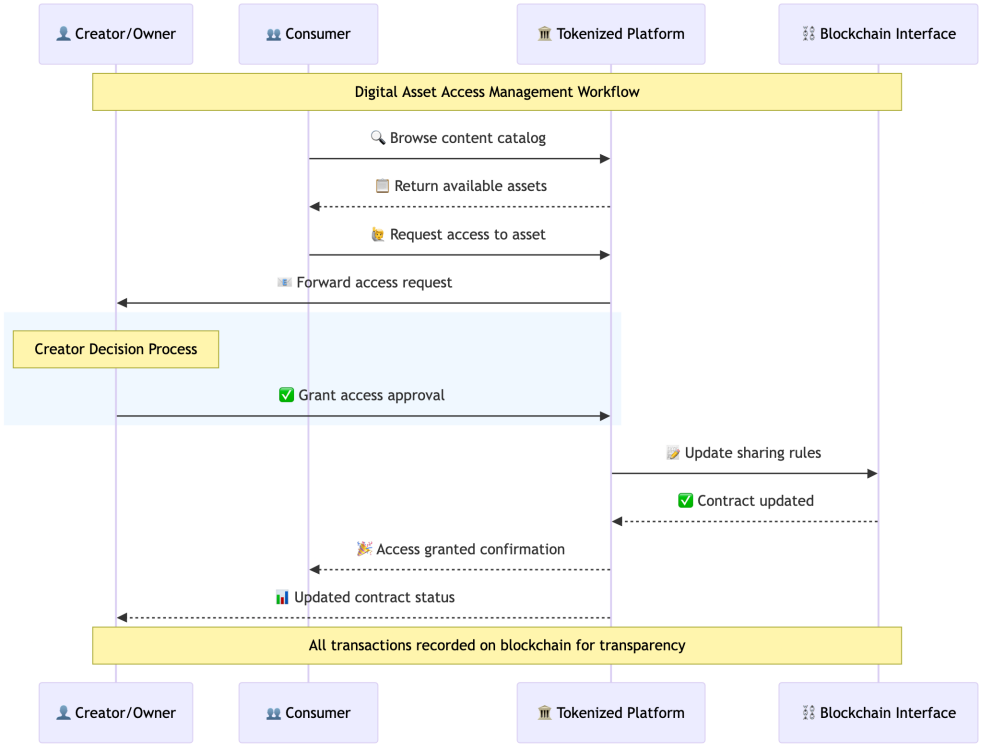


Figure 16.3: Data Access Management in the Tokenized Platform

creators can register new assets, which triggers the automatic generation of a smart contract that links metadata to off-chain content storage. These contracts include fields such as ownership, creation timestamp, and a Content Identifier (CID) pointing to the actual asset stored on IPFS. Three smart contracts are available: Content Ownership Certification, Access and Sharing Management, and NFT Tokenization with Marketplace Integration. The third extends the platform's capabilities by letting users create and trade NFTs directly from the GUI, creating a transparent link between asset identity and its market circulation.

Access Control and Marketplace

The Tokenized Platform's smart contract system manages how users interact with digital assets. Smart contracts control access to tokenized content through a streamlined process. Users can request access to specific content through the Content Catalog, with creators able to accept or decline these requests. The blockchain records all transactions, including access grants, denials, and NFT transfers (see Figure 16.3). When

creators generate NFTs from their tokenized assets, these tokens are registered in smart contracts and appear in a dedicated catalog section. The ERC-721 standard enables NFT creation, transfer, and ownership tracking, providing insights into token ownership and account balances. This creates a verifiable system where users can trade digital assets with automatic ownership transfers.

Web-Based Interface and API Integration

The Tokenized Platform offers all features through a web-based interface (see Figure 16.4). Creators can upload assets, define metadata (license, tags, standards), mint tokens, manage access permissions, and use the NFT marketplace. Developers get the same functionality through a REST API. This allows integration with other SUN components or external services. Key functions include asset listing, tokenization, access requests, and NFT creation or purchase.

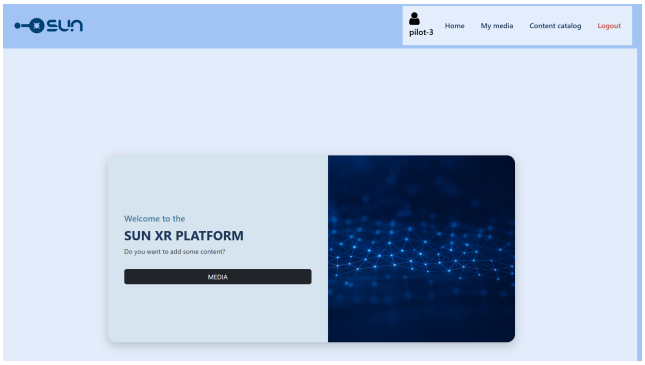


Figure 16.4: Tokenized platfom GUI

Off-Chain Storage and Data Management

The Off-Chain Storage module works alongside blockchain tokenization to handle real-time data. It monitors MQTT topics from OmniBridge and stores incoming data as JSON in MongoDB. Each message gets indexed by session, topic, and component ID. The system automatically generates logs for every stored item. A GUI was implemented to allow complete management of the Off-Chain SUN persistence module, including session and topic management, data storage, logging, and data retrieval. The GUI allows the administrator of the platform to manage in an easy way the configuration of the module, the subscription to specific topics to Omnibridge, and finally offers the possibility to visualize the data stored. These features ensure that the XR session data stays organized, auditable, and reusable.

16.3.3 Cyber Threat Detection

XR technologies create new security challenges. Users become immersed in virtual worlds where they cannot monitor system status or detect security breaches easily. Traditional cybersecurity tools like firewalls and network protection prove insufficient for immersive environments. The Cyber Threat Detection system addresses this gap by providing real-time protection specifically designed for VR and AR applications. The system targets unique attack vectors in XR environments. Malicious actors can manipulate sensory input, alter user awareness, and exploit GPU vulnerabilities to degrade the immersive experience. Unlike traditional attacks that target data theft, XR attacks focus on disrupting visual rendering, causing frame drops, and potentially inducing VR sickness or disorientation in users. More details on the Cyber Threat Detection system will be provided in Chapter 18.

Architecture and Detection Mechanism

The Cyber Threat Detection system operates through a data-driven approach using unsupervised machine learning. It learns normal system behavior patterns and flags deviations as potential threats. The core detection engine employs Isolation Forest[4] algorithms to identify anomalies in real-time performance metrics. The system monitors critical rendering parameters including frame rates, frame durations, and GPU utilization patterns. When attacks target the GPU by injecting malicious texture images or overloading graphical computations, the system detects the resulting performance degradation within seconds. In pilot testing, detection latency averaged0.27 seconds in AR environments and 1.25 seconds in VR environments, with zero false positives.

Alert System and User Protection

When threats are detected, the system triggers multi-modal alerts that combine visual and auditory feedback. The alert design leverages research on immersive notifications to capture user attention effectively. Visual alerts feature bright green backgrounds with yellow warning triangles and exclamation marks, following standard warning iconography. Text size and border strokes ensure readability under diverse environmental conditions. The alerts recommend immediate protective actions such as the removal of the headset or a session pause to prevent user harm from VR sickness, physical disorientation, or psychological stress during attacks. User testing shows high comprehension rates, with participants rating alerts as clear, timely, and actionable in all pilot scenarios.

[4] https://en.wikipedia.org/wiki/Isolation_forest

Implementation and Testing

The system integrates with Unity 3D environments through custom C# scripts and leverages OpenXR plugins for cross-platform compatibility. The testing infrastructure includes HoloLens 2 devices for AR scenarios and Meta Quest 2 systems for VR environments, connected via 5G networks to ensure realistic deployment conditions.

During tests, GPU-based attack simulations used OpenGL API malware that loaded textured images at 5-second intervals, systematically overwhelming the GPU and causing frame delays. The system consistently detected these attacks with an accuracy of >90% while maintaining detection latency below 6 seconds across all scenarios tested.

Security Standards

The Cyber Threat Detection system follows MITRE ATT&CK framework standards, specifically T1499.004 for Application Resource Exhaustion attacks[5]. It aligns with NIST AI Risk Management Framework requirements for the "Detect" and "Mitigate" functions.

16.4 Deployment and Delivery

The SUN Integrated Platform transforms complex XR deployment challenges into simple, manageable processes. Teams can choose their preferred deployment approach based on specific requirements and infrastructure constraints.

16.4.1 Flexible Deployment Options

The platform supports *three* deployment scenarios to meet diverse operational needs:

- *Cloud-hosted environments*: offer immediate access through a development cluster. This option eliminates setup overhead and provides instant access to the platform for testing and development.
- *Self-hosted deployments*: leverage containerization technology to ensure consistent performance across different environments. The Docker Compose setup enables local deployment without internet connectivity—critical for scenarios requiring complete data sovereignty or working in restricted network environments.

[5] https://attack.mitre.org/techniques/T1499/004/

- *Hybrid configurations*: combine local and remote components through experimental MQTT bridging capabilities. This approach allows distributed teams to maintain local autonomy while enabling seamless communication with central systems.

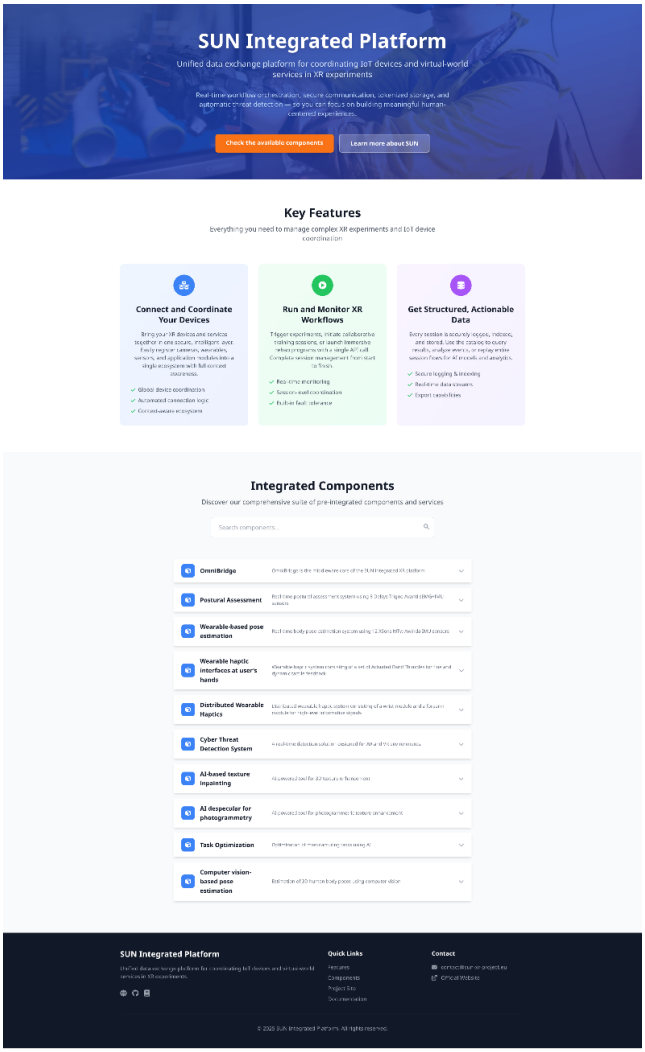


Figure 16.5: Landing page of the SUN Integrated Platform

16.4.2 Streamlined Integration Stack

Early pilot testing posed a challenge for deployment simplicity. Teams struggled with complex multi-component installations that required extensive coordination between different platform elements. The entire Integrated Platform stack deploys through a single Docker Compose setup. This includes the OmniBridge service, MQTT broker, Observability UI, and the Tokenized Platform component. This push-button solution eliminates the fragmented deployment process. Teams do not need to manually coordinate separate installations for each platform component. A single command launches the complete integration environment, whether locally or on-premises.

16.4.3 Per-Part Deployment

Each platform part has its own deployment method, whether through containers or other means. Teams can deploy each one separately rather than using the complete integrated stack. This flexibility proves valuable when projects require only specific platform capabilities or when teams need to integrate individual components into existing infrastructure. The modular approach aligns with the pilots' requirements, where some teams needed local or hybrid deployments while others integrated only selected components. The ability to choose between full-stack and per-part deployments ensures that the platform adapts to diverse environments without unnecessary overhead.

16.5 Conclusions

A dedicated landing page for the SUN Integrated Platform has been created: https://sun.in-two.com/ (see Figure Figure 16.5). It provides external stakeholders with an overview of the different components and capabilities of the SUN Integrated Platform. The integrated SUN platform can support a diverse set of functionalities, ranging from XR content tokenization and off-chain data persistence to immersive interaction and cybersecurity. The modular deployment, enhanced middleware, and comprehensive interface layers position the platform for real-world applications in varied XR environments.

17. Tokenized Platform for Customers Digital Assets Exchange

Ferdinando Bosco[1] *and Vincenzo Croce*[1]

[1]Engineering Ingegneria Informatica S.p.a. (ENG), Italy

Abstract. The Tokenized Platform is a core architectural component of the SUN Platform, designed to manage digital assets and ensure secure, transparent transactions through blockchain technology. By leveraging both fungible and Non-Fungible Tokens (NFTs), the platform facilitates asset certification, tokenization, and access control. It integrates smart contracts to record transactions in a decentralized manner, enhancing trust and traceability. Beyond tokenization, the platform serves as a central data repository, implementing off-chain storage and contributing to the development of a European Dataspace for digital assets. Its multi-layer architecture includes an advanced User Interface (UI), business logic based on Smart Contracts, a Blockchain layer for supporting Ethereum and Polygon blockchains, and a storage layer which utilizes IPFS (InterPlanetary File System) and MongoDB for asset and raw data storage, respectively. The Tokenized Platform is relying on ChainPro, a Web3 innovative modular framework, developed as a research outcome, that supports advanced blockchain technologies. It provides support to facilitate the creation of Web3 applications, enable an easy and fast interaction with the main Layer 1 (Ethereum) and Layer 2 (Polygon) blockchains, and provide a layer of standard interfaces based on open APIs for the implementation of core functionalities. This chapter explores the technological foundations, architectural design, and operational capabilities of the Tokenized Platform. It highlights its role in enabling secure Extended Reality (XR) digital asset management and data sovereignty, as well as its potential impact.

17.1 Introduction

The digital transformation of industries has led to an increasing need for secure, transparent, and interoperable platforms for managing digital assets. The Tokenized Platform, developed within the SUN project, addresses this need by integrating blockchain technology with advanced data management capabilities. It enables the creation, certification, and transfer of digital assets using tokenization mechanisms, while also serving as a persistent data repository for the broader SUN ecosystem.

This chapter provides a comprehensive overview of the Tokenized Platform, detailing its architecture, functionalities, and integration mechanisms. It also discusses the underlying blockchain technologies and tokenization strategies that empower the platform's capabilities.

17.2 Blockchain and Tokenization Mechanisms

Blockchain is a decentralized ledger technology that records transactions across a distributed network of nodes. It ensures immutability, transparency, and security, making it ideal for applications involving digital asset management.

Tokenization is the process of converting rights to an asset into a digital token on a blockchain. The platform supports fungible tokens and Non-Fungible Tokens (NFTs). Smart contracts govern the creation, transfer, and access control of these tokens, ensuring secure and automated asset management [Safitri et al. 2025].

The Tokenized Platform supports two kinds of blockchain infrastructures: Layer 1 (Ethereum) and Layer 2 (Polygon). Ethereum and Polygon blockchains both support the implementation of Smart Contracts as well as the creation of NFTs, the main goals of the Tokenized Platform [Onwubiko et al. 2023].

When Smart Contracts are deployed or executed, the transaction must be processed by a validator. This process requires energy, and using the blockchain implies computation, which users pay for through a mechanism called a gas fee. There is a strict relationship between gas fees and demand; therefore, the greater the demand, the higher the gas fees, and vice versa. Additionally, gas fees are higher during peak periods of computation when the chain has high utilization; therefore, it is advisable to deploy Smart Contracts or create NFTs during the less productive hours of the day when transaction/gas fees are lower. Over time, there have been congestion issues due to limitations of block capacity on the Ethereum blockchain. These problems make gas fees expensive. This is where Polygon comes in to provide a solution to the Ethereum

cost problems. Polygon aims to make it easier for people to access and use Ethereum by providing a scaled-up version of the Ethereum network with lower fees and faster transaction times. While there may be uncertainty about Polygon's role in the future once Ethereum completes its upgrades, Polygon currently serves a valuable purpose in the cryptocurrency ecosystem by helping bring more people in by providing scalable solutions for web3 [Song et al. 2024].

In the Tokenized Platform, the XR Content Creator is able to import digital assets, certify them over the blockchain, define data access management rules, and create NFTs on Polygon or Ethereum, depending on their own specific needs and priorities. Polygon offers low gas fees and a growing market; Ethereum can offer more liquidity and user participation. In the first release of the Tokenized Platform, in the scope of the SUN validation phase with early adopters, a Polygon testnet was selected and integrated into the platform. Testnets are blockchains designed to mimic the operating environment of a mainnet but exist on a separate ledger. These testnets help developers test their applications and smart contracts in a risk-free way before deploying their products to the mainnet environment.

17.3 Architecture

The Tokenized Platform consists of a multi-layer platform able to support the integration of digital assets and raw data with multiple interfaces and communication protocols. The multi-layer architecture, shown in Figure 17.1 includes

- *UI Layer*: Is the entry point of end-users and, in particular, for digital creators that can create and upload digital assets, exploiting the core functionalities based on blockchain technology offered by the tokenized platform: asset certification, tokenization, and access management;
- *Service Layer*: Implements all the business logic of the Tokenized Platform, including services for the digital asset value chain based on Smart Contracts implementation, as well as additional services for raw data storage and logging;
- *Blockchain Interface Layer*: Is the connector layer within the Smart Contracts implemented within the Service Layer and the external blockchain infrastructures. This layer supports the deployment and interaction of smart contracts within two different blockchains: Ethereum and Polygon;

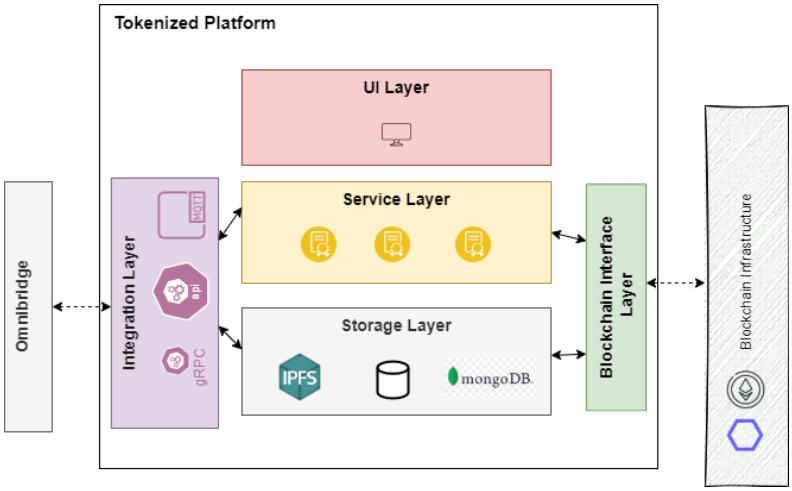


Figure 17.1: Tokenized Platform Architecture.

- *Integration Layer*: This layer implements the interface and communication with external components. In particular, this layer supports several protocols and communication mechanisms, based on the different needs and integration aspects;
- *Storage Layer*: This layer implements the persistence of digital assets and raw data for the entire SUN integrated platform. In particular, two different storage systems are used based on the different data stored: IPFS for digital assets and MongoDB for raw data.

17.4 Main Functionalities

The Tokenized Platform implements two main modules: the *Asset Tokenization*, which allows **digital asset management and certification via blockchain**, and the *Off-Chain Storage*, which is responsible for implementing data **persistence for the overall SUN Integrated Platform** via Omnibridge.

17.4.1 XR Asset Tokenization

Asset Tokenization module includes the tokenization and certification of XR Contents via blockchain and Smart Contracts, implementing three Smart Contracts: asset tokenization with ownership certification data access management, NFT for Tokenization and Marketplace.

Smart Contracts

Smart contracts aim to keep track of all the XR asset interactions expected in the SUN use cases and applications, including the usage of tools and libraries, the creation or adaptation of a media content item, and its provision to the decentralized platform on behalf of the creators, and the access and download on behalf of the consumers. The smart contracts also allow for fine-grained monitoring and management of these actions as well as their translation into tokenized-based values.

To implement all the expected functionalities, the Tokenized Platform implemented three different Smart Contracts:

- XR Content ownership certification (with decentralized storage and fungible tokens creation);
- Digital Asset Management (Access and Sharing rules);
- NFT Tokenization and Marketplace.

The first version of the Tokenized platform already implemented the first two smart contracts, ensuring the possibility to certify XR Contents in a secure and decentralized way, as well as to manage the access and sharing of these XR Contents. The NFT Tokenization mechanism was implemented in the final version of the platform.

XR Content ownership certification Every time new XR content is imported into the Tokenized Platform, a new Smart Contract is created and linked to the imported content. The workflow is shown in Figure 17.2.

The Smart Contract is deployed starting from a predefined one that implements the following attributes:

- Ownership (address of content's owner);
- Creation Time (the timestamp of the content creation in the Blockchain);
- CID (Content Unique Id);
- URL (Link to the real content);
- Asset Tokens (ERC20 Tokens assigned for tokenization).

Smart Contract only stores metadata of the XR Contents while the real asset is stored on a decentralized file system (IPFS), and a unique content id (CID) is created and only available for the content owner.

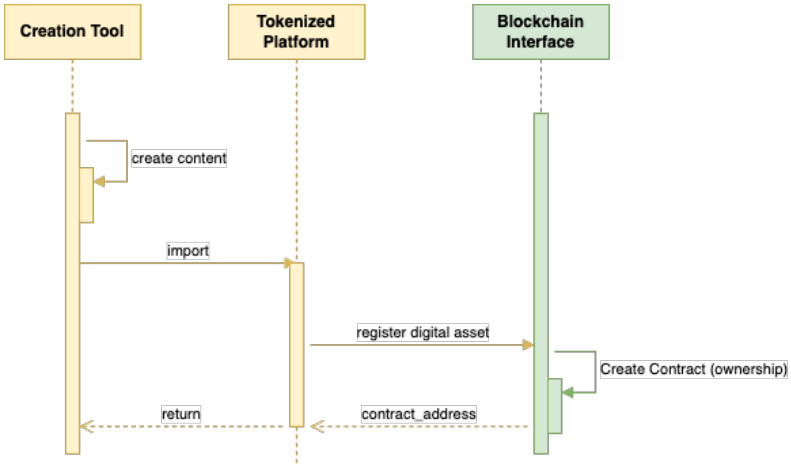


Figure 17.2: Asset Tokenization and Ownership Smart Contract - Sequence Diagram.

Digital Asset Management After the creation of a specific smart contract linked to each content of the SUN Platform, the Creator is able to define a basic method for accessing and sharing models.

In this first release of the Tokenized Platform, the users are able to request access to the specific XR Contents by navigating the Content Catalogue, and the Creator/Owner is able to accept or decline the access request. An example workflow is shown in Figure 17.3.

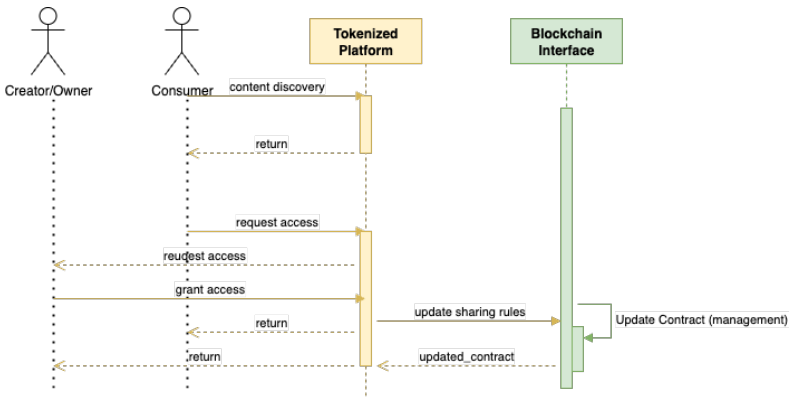


Figure 17.3: Data Access Management Smart Contract - Sequence Diagram.

NFT Tokenization and Marketplace In addition to the basic Data Management feature, the Content Creator is also able to create an NFT token for each specific asset. The NFT (Non-Fungible Token) allows for the identification in a unique way a digital asset by a "token id", as described in ERC-721 standard. The NFT is connected to an asset through a set of metadata. ERC-721 provides an API to:

- Create NFTs;
- Transfer NFTs;
- Gain insights such as:
  - Learn who owns a specific token;
  - Get the supply of such tokens in the network;
  - Find the balance of a specific account in terms of NFTs.

An example workflow is shown in Figure 17.4.

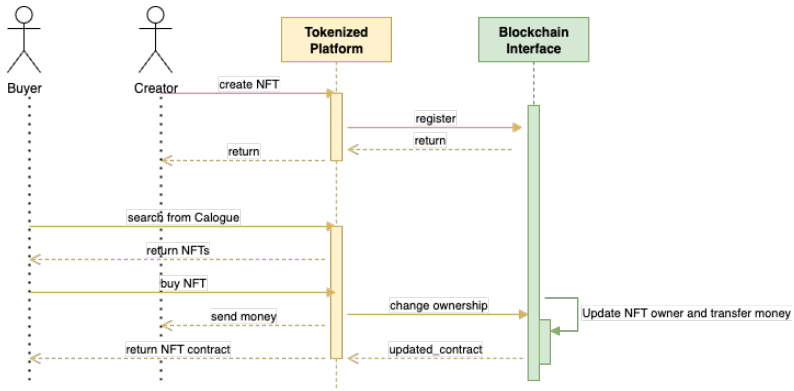


Figure 17.4: NFT Management Smart Contract - Sequence Diagram

Graphical User Interface

All the main features implemented by the Tokenized Platform for digital asset management and tokenization are available both via Graphical User Interface (GUI) in a dedicated web app, and via REST APIs.

All the sections of the GUI with descriptions of the functionalities are reported below.

Registration/Login Creators and end-users are able to register and login to the web app with their own credentials. A unique token is generated for each session to ensure privacy and security throughout navigation.

Asset Tokenization The main section of the GUI allows us to upload any kind of digital asset for content ownership certification. It includes several parameters which can be filled in for each content (e.g., License, Tags, Standards, etc), enriching the content's description and categorization, facilitating at the same time the discovery in the asset catalogue. All these additional parameters are optional, and only the title, description, and media file are mandatory.

My Assets The section provides a list of already uploaded and tokenized assets, including the details about blockchain certification (smart contract address) as well as additional useful info, and the possibility to manage the data access and create NFT (a new feature of the second release).

Asset Catalogue All the uploaded and certified content is available to any user of the Tokenized Platform. The users can discover other creators' content via the Content Catalog sections. In this section, it is possible to view content information and metadata, as well as to request access to specific content. The content becomes available only after access is granted by the Creator

Access Management Consumers can request access to specific content via the Content Catalog. Any request is under the approval of the content Creator. After the confirmation of the Creator, the consumer can access the content itself, and the data transactions are stored in the blockchain.

NFT Creation and Marketplace The Smart Contract for NFT management allows the creation of an NFT starting from a Tokenized Asset. The creator can select their own assets and create NFTs, adding some additional information.
When an NFT is created and registered into a specific Smart Contract on the blockchain, it appears in a dedicated section of the Catalog.
The buyers can discover and buy NFTs from the Catalog. When the purchase is confirmed, the ownership of the NFT is transferred from the original owner to the buyer.

17.4.2 Off-Chain Storage - SUN Platform Persistence

The Tokenized Platform modules dedicated to Off-Chain Storage allow the implementation of the overall SUN platform persistence. The Off-Chain Storage module is fully integrated and complementary to the Omnibridge solution and is able to add a persistence layer, interfacing with all the SUN Components, exploiting the same communication channels offered by Omnibridge.

The module implements a dedicated MQTT handler for interacting as "publisher" and/or "subscriber". In this way, it is able to implement two main functions:

- Collection of raw data and storage in a NoSQL database (MongoDB) using JSON format for metadata;
- Publication of log messages for any data received from the SUN Platform, with specific status messages (success or errors);
- Possibility to configure automatically or manually topics and sessions;
- GUI for the topic, session, and data storage management.

The module includes a dedicated channel for monitoring the broadcast channel in order to allow the platform to subscribe to any new channel and ensure data persistence. Alternatively, the sessions and related topics to be integrated can be configured via a configuration file before the Tokenized Platform installation, or added in a later stage via REST APIs or using the new GUI provided for the component's owners.

The platform is also able to verify the data format for each message within any channel (XML and JSON formats supported). In case the quality check on the data format is correct, the message format is maintained, and the data format is stored as metadata; otherwise, the message is stored as a string.

Below is an example of data stored:

```
{
 "_id": { "674f2e2cee9d7fc8c737a393" ,
 "createdOn": 1733242412554,
 "updateOn": 1733242412554,
 "topic": "sessions/584839f2-d04b-4639-bd0b-c7d1051bffbe/
       7a65d3bc-4f9a-4b70-8248-dd2e10a265b0",
 "componentIdOwner": "7a65d3bc-4f9a-4b70-8248-dd2e10a265b0",
"sessionId": "584839f2-d04b-4639-bd0b-c7d1051bffbe",
 "message": {// message payload},
```

```
  "messageType": "json",
}
```

For any message received by the Tokenized Platform, a log message is published on the topic "*sender_topic*/log" where "*sender_topic*" is the original topic where the message is published and read by the Tokenized Platform. In this way, a new dedicated topic is created for any session, and the log is accessible only by the owner of the message.

The log message includes information about the status of persistence as well as information for retrieving the message in a second phase on the Tokenized Platform.

In fact, all the messages stored within the Tokenized Platform are available and findable by SUN Components and/or actors. In case of JSON messages, the Tokenized Platform also offers a parameterizable query for finding messages based on its internal fields.

Finally, a GUI was implemented to allow complete management of the Off-Chain SUN persistence module, including session and topic management, data storage, logging, and data retrieval.

17.5 Conclusions

The Tokenized Platform represents a significant advancement in digital asset management and data persistence. By integrating blockchain technology with robust data storage and communication mechanisms, it ensures secure, transparent, and scalable operations. Its modular architecture and standardized interfaces make it adaptable to various environments, supporting both cloud and on-premises deployments.

The platform's contribution to the European Dataspace initiative highlights its potential in promoting data sovereignty and interoperability across digital ecosystems. As digital assets become increasingly central to industry and culture, platforms like this will play a crucial role in shaping the future of secure digital interactions.

References

Onwubiko, Austine, Raman Singh, Shahid Awan, Zeeshan Pervez, and Naeem Ramzan (2023). "Enabling trust and security in digital twin management: a blockchain-based approach with ethereum and ipfs". In: *Sensors* 23.14, p. 6641.

Safitri, Nurdiana, Andry Alamsyah, and Dian Puteri Ramadhani (2025). "Comparative Network Dynamics of Ethereum and Polygon: Insights into NFT Ecosystem Structures and Behaviors". In: *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*. IEEE, pp. 1–6.

Song, Han, Zhongche Qu, and Yihao Wei (2024). "Advancing blockchain scalability: An introduction to layer 1 and layer 2 solutions". In: *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*. IEEE, pp. 71–76.

18. Cyber Threat Detection in XR

Junyi Zou[1], *Georgios Loukas*[1], *Riccardo Bovo*[1], *and Ali Hamza*[1],

[1] University of Greenwich (UoG), UK

Abstract. This chapter presents a Cyber Threat Detection tool for Virtual Reality (VR) and Augmented Reality (AR) technologies, addressing the critical challenge of technological safety and user protection. Our investigation has revealed that traditional cybersecurity mechanisms, which have historically focused on protecting digital devices through firewalls, authentication, and network security, are insufficient for immersive technologies. VR and AR technologies capture different types of personal data, including physical movements, physiological responses, spatial awareness, and potentially unconscious behavioural patterns. The immersive nature of these technologies significantly compromises users' ability to monitor system status or recognize potential security breaches. In this chapter, we show our work towards an intrusion detection system capable of warning users about cybersecurity threats during immersive experiences. The system achieved high detection accuracy and low detection latency. It was tested across multiple pilots, confirming its practical readiness and reliability in real-world XR scenarios.

18.1 Introduction

The growing use of XR technologies in diverse application domains intensifies the importance of safeguarding user data and system integrity. The digital nature of VR and AR creates unique vulnerability points where cyber-attacks can manipulate sensory stimulation, alter user awareness, and compromise targeted experiences. Potential risks include not only privacy breaches but also disruptions to user physical safety, particularly in critical use cases such as remote surgery or rehabilitation.

Traditional security measures, such as firewalls and endpoint protection, are inadequate in addressing these threats. Unlike conventional computing environments, XR

platforms continuously collect and react to highly sensitive, real-time behavioural data, such as gaze, motion, and voice input. This introduces novel attack surfaces, including motion-based inference and sensory distortion attacks.

Existing research has explored various aspects of these vulnerabilities. For instance, [Luo et al. 2024] developed Heimdall, a system capable of inferring typed input from controller sounds with high accuracy. [Meteriz-Yildiran et al. 2022] used hand trajectory data for text inference, while [Al Arafat et al. 2021] demonstrated keystroke recognition using Wi-Fi CSI. Similarly, [Ling et al. 2019] and [Shi et al. 2021] have shown how accelerometers and motion sensors can be exploited to infer passwords or eavesdrop on conversations. In physical manipulation attacks, [Tseng et al. 2022] demonstrated how perceptual interference can lead users to perform unsafe movements.

Broader taxonomies and threat models have been proposed to formalise these risks. [Casey et al. 2019] introduced a taxonomy of VR-specific attacks, including the human joystick attack, in which users are covertly guided to physical obstacles, and overlay manipulations that inject unauthorized visual content. A 3-dimensional privacy risk model tailored to XR was proposed by [Yamakami 2020], while [Qamar et al. 2023] published a systematic taxonomy of XR threats to support the development of detection frameworks. Expanding on this, [Gulhane et al. 2019] assessed security and privacy risks in VR learning environments, emphasizing the limitations of current safeguards in preventing physically dangerous scenarios.

Meanwhile, biometric and behavioural data have been studied both as vulnerabilities and as tools for authentication. [Pfeuffer et al. 2019] and [Kupin et al. 2019] utilized behavioural biometrics in VR to authenticate users based on unique motion signatures. [LaRubbio et al. 2022] proposed gaze-based authentication mechanisms integrated into VR workflows to enhance security in occupational settings. More recently, [Yang et al. 2023] highlighted the risk of mid-air gesture leakage, showing how simple RGB cameras and machine learning could decode passwords from hand movements.

These studies demonstrate that XR systems face complex, multi-dimensional cyber risks. However, defence strategies remain in their infancy. Most approaches focus on prevention, primarily via authentication and secure data handling, rather than on real-time detection or user-centered feedback during attacks.

[Happa et al. 2019] developed a theoretical framework of XR attack surfaces, while [Odeleye et al. 2021] presented a system to detect frame-rate manipulation. Our contribution builds on this work by providing a fully developed, real-time intrusion detection system applicable across both VR and AR platforms. In doing so, we seek to fill an important gap in XR security research: making the transition from abstract threat modeling to user-facing detection and intervention tools.

18.2 Methodology and Results

18.2.1 Background and Design Motivation

As Extended Reality (XR) technologies, ranging from healthcare to industrial training, become more integrated into daily activities, their exposure to cyber threats increases. However, conventional cybersecurity strategies such as firewalls and endpoint protection are not sufficient for immersive environments. XR systems continuously capture rich streams of behavioural data, including gaze, gesture, voice, and motion, creating new attack surfaces and diminishing the user's capacity to detect anomalies. Furthermore, immersive users are often physically engaged and perceptually overloaded, reducing their ability to notice or interpret traditional cybersecurity indicators like system slowdowns or graphical glitches. This necessitates an approach that not only detects threats but also communicates them in a clear, timely, and non-intrusive manner. The work presented here introduces a real-time Cyber Threat Detection system, designed to operate seamlessly within XR environments and provide user-facing alerts when suspicious activity is identified.

18.2.2 System Overview

The Cyber Threat Detection system is built on an unsupervised machine learning framework, specifically designed to operate within XR environments. At its core is the Isolation Forest algorithm, a well-established anomaly detection method capable of identifying outliers in system behaviour without requiring labelled training data. This approach is particularly well-suited for XR, where threats may emerge unpredictably and lack predefined signatures.

Isolation Forest works on a simple principle: anomalies are easier to isolate than normal data points. The algorithm builds random binary trees by making random splits in performance data. Normal XR metrics cluster together—frame rates around 90 FPS, GPU usage near 65%. These clustered points need many splits to separate from each other. A GPU attack drops frame rates to 30 FPS and spikes GPU usage to 95%. This outlier gets isolated quickly with just one or two splits. The algorithm measures path length—how many splits it takes to isolate each data point. Short paths indicate anomalies and trigger security alerts.

The data is sent in real time to a cloud-based detection module, where the anomaly detection algorithm evaluates it for signs of potential cyberattacks.

The prediction results from the server are then sent back to the main XR experience (named SunProject), which integrates the detection system as a modular Unity package. If an anomaly is detected, SunProject delivers a user-facing alert via visual and auditory cues, allowing the user to take appropriate action. This architecture ensures minimal impact on local performance while enabling rapid detection and response.

To simulate and test attacks, a separate Unity project called TestAttack is used to trigger GPU-based denial-of-service behaviour. This project injects high-load graphical operations designed to disrupt performance, enabling controlled testing of detection accuracy and system responsiveness (see Figure Figure 18.1).

The modular and cloud-integrated nature of this system allows it to scale easily across XR applications while maintaining low latency and high detection accuracy.

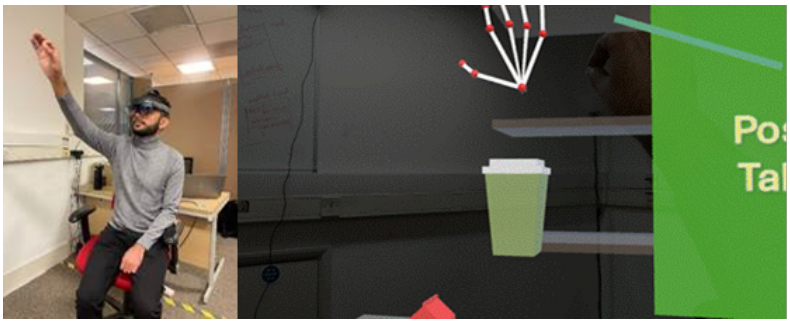


Figure 18.1: Participant who has been affected by the attack and is trying to place the object in the right position.

18.2.3 XR Environment Integration

The detection system is designed to integrate with both VR and AR platforms. In VR, it was implemented within Unity applications running on Meta Quest 3, using the XR Interaction Toolkit and OpenXR pipeline. For AR, it was deployed on HoloLens 2, using Microsoft's Mixed Reality Toolkit (MRTK). These environments feature immersive tasks including object manipulation, precision interactions, and collaborative sessions. The module collects telemetry such as rendering statistics, user interaction latency, and system performance indicators.

Alerts are a critical component of the proposed Cyber Threat Detection system, designed to prompt timely user action in response to potential security threats. Drawing

from established human-computer interaction (HCI) principles and prior research on immersive notifications, the alert system uses a multimodal communication approach to ensure that alerts are noticeable, comprehensible, and actionable within immersive environments. The alert system combined both visual and auditory warning.

The use of combined modalities is grounded in findings that auditory icons paired with visual stimuli reduce cognitive load and enhance user performance. This multimodal strategy ensures fast user reactions, clearer understanding of threat severity, and more effective responses sensitive or high-stress scenarios.

Visual alerts are designed with high contrast, motion, and iconography to immediately capture attention. Text size has been refined through iterative user testing to ensure readability, while border strokes enhance legibility against complex backgrounds. A bright green background is used to draw general attention, while a yellow triangle with a bold exclamation mark signals urgency, following conventional warning standards. These design elements prioritize visibility and clarity across varying environmental conditions.

Auditory feedback consists of spatial sound cues and verbal instructions, selected to complement visual information without overwhelming the user. The audio components are carefully chosen to align with cognitive load considerations and ensure accessibility. Together, the visual and auditory channels work in tandem to deliver alerts that are intuitive, effective, and respectful of the user's immersive experience (see Figure Figure 18.2).

18.2.4 Attack Simulation and Detection Mechanism

To evaluate the robustness and responsiveness of the proposed detection system, a targeted cyberattack was designed to simulate a GPU-based Denial-of-Service (DoS) threat within XR environments. This synthetic attack aims to mimic real-world threat behaviours that compromise rendering performance, a particularly critical vector in immersive systems where perceptual continuity directly impacts user safety and comfort.

The simulated attack operates by injecting large, high-resolution textures into the graphics rendering pipeline at controlled intervals. These textures are intentionally oversized and memory-intensive, causing a spike in GPU workload that overwhelms the rendering engine, leading to frame drops, increased latency, and perceptual disruptions such as jitter, screen freezing, and ghosting effects. Even short-lived rendering disruptions in XR can induce motion sickness, reduce interaction fidelity, and impair spatial awareness, posing both a usability and safety concern.

To maintain the integrity of the main system, this synthetic malware is executed in a separate Unity project named TestAttack, rather than being embedded in the main

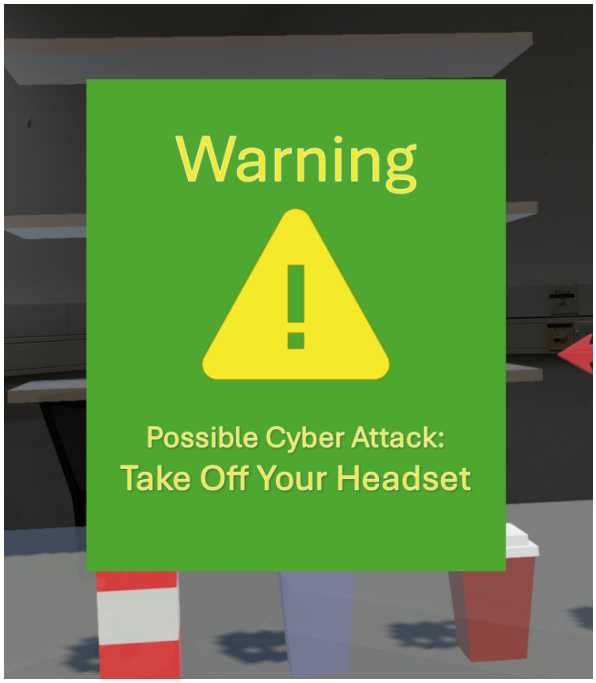


Figure 18.2: Design of warning signs in Cyber Threat Detection system

experience. This approach ensures that the malware remains isolated from production builds, preventing unintended side effects during regular development while also enabling focused evaluation of the detection system under controlled attack conditions.

The implementation of the attack exploits CPU-based injection of rendering calls, executed as a lightweight background process that mimics stealthy threat behaviour. It does not cause the application to crash but introduces performance anomalies that are often difficult for users to diagnose, making it a subtle but effective test case for anomaly detection.

The attack conforms to the MITRE ATT&CK framework, specifically subcategory T1499.004: Application or System Resource Exhaustion, which outlines threat behaviours that intentionally degrade system performance by exhausting computational resources. By aligning with this framework, the evaluation ensures methodological consistency with recognized cybersecurity standards.

During runtime, the detection module continuously collects telemetry (including frame time variance, interaction latency, and scene render time) via the FrameLogger component. This telemetry is transmitted to a cloud-hosted detection model, where an Isolation Forest algorithm processes the data to identify anomalous patterns. Once an

anomaly is detected, the prediction is sent back to the main XR application (SunProject), which integrates the detection system as a modular Unity package. A user-facing alert is then triggered, using multimodal cues to prompt immediate action, such as pausing the experience or removing the headset.

This cloud-supported architecture allows the detection model to scale across devices and XR platforms while minimizing client-side computational load. It also provides flexibility for future adaptation, such as expanding the threat model to include network-based intrusions, spoofing attacks, or data manipulation scenarios relevant to collaborative XR settings.

18.2.5 Evaluation and Observed Outcomes

Across all testing scenarios, the anomaly-based intrusion detection system demonstrated a high degree of accuracy. Detection accuracy consistently exceeded 90%, indicating that the model was reliably able to differentiate between normal and anomalous XR system behavior. This reinforces the suitability of Isolation Forest in high-dimensional, real-time monitoring contexts such as immersive environments, where data patterns evolve continuously during user interaction.

Latency measurements showed that the system could issue alerts in under 1.25 seconds following the onset of GPU-based cyber attacks. This prompt detection is critical in immersive experiences, where even brief disruptions in system responsiveness can degrade user presence and lead to disorientation. Low latency ensures that users receive timely warnings and can take precautionary action before the attack fully impacts performance.

Throughout all evaluation conditions, the model recorded zero false positives. This is particularly important in XR settings where excessive or unwarranted alerts can erode user trust, contribute to cognitive fatigue, or detract from task performance. The system's precision reflects careful threshold tuning and the robustness of the model under varying performance profiles.

User comprehension of alerts was assessed using the Intrusion Detection System Questionnaire (IDSQ), a tool developed within the SUN project to assess how well users understand system prompts in immersive environments. Participants reported that alerts were clear, contextually appropriate, and easy to act upon, highlighting the effectiveness of the multimodal communication strategy and the importance of intuitive design in security UI.

Emotional responses to the alerts were evaluated using the Self-Assessment Manikin (SAM). Participants typically reported high levels of arousal, suggesting that alerts were

successful in drawing attention. However, valence scores remained neutral to positive, indicating that users did not experience fear or distress. This balance between urgency and comfort is essential in XR environments, where emotional overload can reduce task performance or induce sickness.

Analysis of user interaction logs and post-session feedback revealed that both visual and auditory cues were instrumental in conveying alerts. Some users responded immediately to high-contrast visual overlays and warning icons, while others were more reactive to spatialized audio cues or verbal prompts. The redundancy in modalities ensured that critical information reached users through at least one perceptual channel, reinforcing the benefits of a multimodal approach.

18.3 Conclusions

As XR technologies continue to expand across critical sectors such as healthcare, industrial training, and remote collaboration, ensuring their cybersecurity becomes an increasingly urgent priority. This chapter presented the design, implementation, and evaluation of a real-time Cyber Threat Detection system tailored for both VR and AR environments. By leveraging GPU-based attack simulations, immersive testbeds, and multimodal alert mechanisms, we demonstrated the feasibility and effectiveness of user-facing threat detection tools in XR. Across three pilot deployments in different countries and use cases, the system achieved high detection accuracy, low latency, and strong user comprehension with zero false positives.

Our results reinforce the importance of combining technical threat detection with human-centered design. Multimodal alerts, combining visual and auditory cues, proved to be effective in prompting user responses without inducing discomfort. Furthermore, packaging the detection system as a Unity Package Manager (UPM) module allowed for scalable, cross-platform deployment, supporting adoption across varied XR settings.

Future work will focus on expanding the scope of detectable attack vectors, refining alert personalization based on user behavior, and integrating feedback loops that adapt detection strategies over time. Ultimately, our goal is to contribute to a safer, more resilient XR ecosystem where immersive experiences can be both innovative and secure.

References

Al Arafat, M. Y., M. M. Ahmed, M. Al Hasan, M. A. Rahman, and M. M. Rahman (2021). "VR-Spy: Towards Inferring Typed Input on Virtual Keyboards in Virtual Reality Headsets Using WiFi". In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 3303–3305.

Casey, P., C. Chalmers, and T. Ha (2019). "A Taxonomy of VR-Specific Threats and Attacks". In: *Proceedings of the IEEE Conference on Games (CoG)*, pp. 1–8.

Gulhane, A., A. Vyas, R. Mitra, R. Oruche, G. Hoefer, S. Valluripally, P. Calyam, and K. A. Hoque (2019). "Security, Privacy and Safety Risk Assessment for Virtual Reality Learning Environment Applications". In: *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, pp. 1–9.

Happa, J., A. Williams, and S. Smith (2019). "Extended Reality (XR) Attack Surfaces: A Security and Privacy Threat Analysis". In: *Proceedings of the International Conference on Cyberworlds (CW)*, pp. 102–109.

Kupin, A., B. Moeller, Y. Jiang, N. K. Banerjee, and S. Banerjee (2019). "Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments". In: *MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I*. Vol. 25. Springer, pp. 55–67.

LaRubbio, K., J. Wright, B. David-John, A. Enqvist, and E. Jain (2022). "Who Do You Look Like?–Gaze-Based Authentication for Workers in VR". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 744–745.

Ling, X., Q. Xu, T. Wu, and K. Zhang (2019). "A Study on Sensor-Based Password Inference Attacks in Virtual Reality". In: *Computers & Security* 87, p. 101571.

Luo, B., S. Liu, M. Fereidouni, K. Zhang, and H. Chen (2024). "Heimdall: Keystroke Inference in Virtual Reality Using Acoustic Side Channels". In: *Proceedings of the IEEE Symposium on Security and Privacy (SP)*.

Meteriz-Yildiran, C., M. M. Ahmed, D. Iorga, A. Tuncer, and H. A. Varol (2022). "Predicting Hand-Tracked Air-Tapping Actions in VR Using Convolutional Neural Networks". In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4.4, pp. 445–456.

Odeleye, O., Y. Liu, G. Loukas, and M. Yampolskiy (2021). "Towards Real-Time Intrusion Detection in VR: A Case Study on Frame-Rate Attacks". In: *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 525–526.

Pfeuffer, K., M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt (2019). "Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.

Qamar, S., Z. Anwar, and M. Afzal (2023). "A Systematic Threat Analysis and Defense Strategies for the Metaverse and Extended Reality Systems". In: *Computers & Security*, p. 103127.

Shi, Y., Q. Zhao, Y. Yang, and Y. Liu (2021). "Eavesdropping on Speech in Virtual Reality Using Motion Sensors". In: *IEEE Transactions on Dependable and Secure Computing* 18.3, pp. 1185–1197.

Tseng, E. et al. (2022). "The Risks of Misperception in VR: Virtual Physical Perceptual Manipulations (VPPMs)". In: *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–10.

Yamakami, T. (2020). "A Privacy Threat Model in XR Applications". In: *Advances in Internet, Data and Web Technologies*. Ed. by L. Barolli, Y. Okada, and F. Amato. Cham: Springer International Publishing, pp. 384–394.

Yang, W., X. Dengxiong, X. Wang, Y. Hu, and Y. Zhang (2023). ""I Can See Your Password": A Case Study About Cybersecurity Risks in Mid-Air Interactions of Mixed Reality-Based Smart Manufacturing Applications". In: *Journal of Computing and Information Science in Engineering*, pp. 1–12.

Human-Centered XR Scenarios and Real-World Case Studies

In the rapidly evolving landscape of science and innovation, technology alone cannot define progress. This section invites us to reimagine research and innovation through the lens of human experience, where individuals and their environments are not external beneficiaries but active co-creators of transformation. By focusing on human values, cultural contexts, and real-world needs, this perspective moves beyond functional design to embrace empathy, inclusion, and ethical foresight. Within this context, "scenarios" serve as living laboratories: frameworks through which disciplines can converge to anticipate societal challenges and test future visions in tangible, participatory settings. Whether addressing digital transformation, health, or industry, each scenario grounds technological advancement in authentic human narratives. This alignment between imagination and reality enables us to explore not only what technology can do, but what it should do and for whom. This section underscores SUN's core mission: fostering a new paradigm of innovation that is not only intelligent and efficient but also meaningful, equitable, and deeply human.

19. Humanity and Ethics Driven Scenarios

Fabio Perossini[1], *Giuseppe Caracciolo*[1], *Silvia Boi*[1], *Devin Bayer*[2], *Michaela Wiese*[2], *and Cong Yao*[3]

[1] KPeople Research Foundation (KPRF), Malta
[2] Outdoor Against Cancer (OAC) Germany
[3] Vrije Universiteit Brussel (VUB), Belgium

Abstract. This Chapter outlines how the SUN project integrates Extended Reality (XR) and Artificial Intelligence (AI) within a human-centred and ethically responsible framework. The approach is based on participatory design, ensuring that patients, clinicians, workers, and technology developers contribute to shaping solutions that respond to real needs while respecting dignity, inclusivity, and trust. It describes three main pilot scenarios. The first pilot focuses on rehabilitation, where patients recovering from musculoskeletal injuries engage in immersive, gamified XR therapy supported by real-time monitoring and adaptive feedback. The second pilot addresses workplace safety and collaboration, using XR for object tracking, Personal Protective Equipment (PPE) detection, and augmented workflows to improve awareness, reduce risks, and enhance cooperation. The third pilot explores autonomy for people with disabilities, leveraging multimodal interfaces such as eye tracking and electromyography (EMG) sensors to foster communication, independence, and social participation. Across these scenarios, the emphasis is not only on technical performance but also on ethical safeguards, accessibility, and user well-being. This chapter highlights challenges such as ergonomics, cognitive load, empathy, and trust, framing XR as both a technological innovation and a socio-cultural enabler that requires careful alignment between innovation, ethical principles, and human values.

19.1 Introduction

The SUN (Social and hUman-ceNtered XR) project emerges at the intersection of cutting-edge Extended Reality (XR), Artificial Intelligence (AI), and a human-centred design philosophy to create socially meaningful and ethically grounded digital experiences. At its core, SUN accepts that the immersive potential of XR—spanning Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR)—is not only a technical innovation but also a socio-cultural one. Its ambition is to transition from isolated XR applications to a new paradigm where immersive environments are seamlessly integrated into the physical world, deeply embedded with human and social values[1].

Historically, XR technologies have made significant progress in industrial training, gaming, and simulation. However, most implementations have remained isolated from real-world interaction or constrained to visual overlays, lacking deeper semantic or physical integration. SUN seeks to overcome these limitations by introducing adaptive, AI-powered XR systems capable of learning from real-world interactions and enabling seamless, bi-directional feedback between physical and virtual domains. This involves real-time 3D acquisition, sensor-based physical modelling, and embodied AI [SUN D2.2 2024] to support personalized, context-aware interactions across domains such as health rehabilitation, workplace safety, and disability inclusion.

19.2 Co-Designing Pilot Scenarios: Methodology and Vision

Designing impactful and ethically grounded pilot scenarios in human-centred XR systems requires a multidisciplinary approach that integrates technological innovation with user needs, cognitive models, and societal values. Within the SUN project, the co-design methodology reflects a deep commitment to responsible innovation. This chapter outlines the processes, rationale, and structure behind the pilot scenarios developed across the three use cases—rehabilitation, industrial safety, and assistive interaction for people with severe disabilities.

19.2.1 User-Centred Scenario Development in SUN

The SUN project adopts a scenario-based design approach rooted in participatory, human-centred principles. This methodology is critical to ensuring that the XR tech-

[1]European Commission, (2022). SUN Grant Agreement No. 101092612. Horizon Europe Programme

nologies developed are not only functional but also socially acceptable, safe, and meaningful. In SUN, pilot scenarios are conceived as structured environments that simulate real-world challenges, enabling iterative testing of XR solutions in rehabilitation, industrial safety, and accessibility. Each scenario is designed to evaluate user performance, acceptability, and engagement while collecting multimodal data for further analysis.

The co-design process begins with user requirement elicitation (as reported in [SUN D2.2 2024]), followed by iterative prototyping and validation. Stakeholders, including patients, workers, clinicians, therapists, and technology providers, are involved throughout, ensuring the pilots remain aligned with user expectations and ethical standards. A dedicated planning framework [SUN D6.1 2024] supports this process by outlining technical, organizational, and regulatory constraints for each scenario.

19.2.2 Pilot Use Cases: Rehabilitation, Workplace Safety, and Severe Disabilities

The SUN pilots are anchored in concrete, high-impact use cases.

Rehabilitation (Pilot 1): Patients recovering from orthopaedic injuries or surgeries engage with XR environments that simulate daily tasks and therapeutic exercises. Gamification, real-time feedback, and avatar-guided movement are used to enhance motivation and track recovery metrics such as range of motion and posture alignment.

Workplace Safety and Interaction (Pilot 2): Industrial workers navigate a digitally augmented shopfloor environment equipped with object tracking, Personal Protective Equipment (PPE) detection, and task prioritization systems. The scenario tests how XR can reduce risk perception gaps, promote safety protocols, and support collaborative workflows.

Interaction and Autonomy for Disabilities (Pilot 3): Individuals with motor or verbal impairments interact with XR applications using multimodal interfaces, such as eye tracking or Electromyography (EMG) sensors, designed to restore autonomy, emotional expression, and social connection. Virtual avatars act as embodied proxies, helping users to participate in shared environments and communication loops.

Each use case is governed by a structured protocol [SUN D6.1 2024] that defines objectives, evaluation metrics, duration, participant roles, and ethical safeguards.

19.2.3 Embedding Human Vision in Scenario Planning

The SUN project wanted to ensure that scenario design did not just incorporate a purely technical dimension but also reflected ethical and social concerns raised by patients and

their communities. This broader human-centric vision required balancing clinical goals, technical feasibility, and the social realities of those who will eventually use XR systems. In practice, this meant raising issues such as accessibility, fatigue management for cancer survivors, trust in data handling, and the importance of emotional engagement and empathy in rehabilitation scenarios.

Beyond technology validation, SUN pilots serve as probes into the dynamics of human vision, cognition, and embodied perception in XR environments. The co-design process intentionally integrates findings from cognitive neuroscience, User eXperience (UX), and perceptual psychology to enhance the naturalness and reliability of the interactions. This includes: i) optimizing visual saliency to guide attention through adaptive cues; ii) ensuring motion congruency between physical and virtual feedback; iii) designing avatars and interfaces that reflect social cues, fostering empathy and trust; iv) accounting for cognitive load and user fatigue in prolonged sessions, particularly in rehabilitation and industrial contexts.

A human vision for SUN is not only a matter of technical optimisation but also about a conscious alignment with human perception, cognitive comfort, and social dimensions. While patient associations were not directly involved in the technical engineering of avatars, motion congruency, etc., it played a role within the consortium by ensuring that human factors were discussed early in scenario design meetings. This contribution helped to keep user well-being in focus, even in early, abstract phases of development. This approach emphasized the importance of designing immersive experiences that feel intuitive, empathetic, and emotionally supportive, especially important for vulnerable users in the context of rehabilitation.

19.3 Human Vision and Perception in Extended Reality

19.3.1 From Visual Ergonomics to Perceptual Feedback

XR applications introduce unique visual challenges due to their immersive nature, including vergence, accommodation conflict, field-of-view constraints, motion sickness, and latency-induced discomfort [SUN D2.2 2024]. These ergonomic issues, if not addressed, can diminish user engagement, limit usage time, and reduce therapeutic or operational effectiveness, especially in vulnerable populations such as patients undergoing rehabilitation or users with motor or cognitive impairments. For this reason, the SUN platform integrates real-time monitoring of gaze, posture, and head orientation using wearable sensors and computer vision to minimize strain and dynamically adapt visual stimuli. Visual ergonomics also intersects with task-specific needs: for example,

rehabilitation tasks may require high-contrast visual targets in specific regions of the visual field, while industrial safety scenarios may prioritize peripheral awareness and depth cues to detect moving hazards. In SUN, visual parameters (e.g., brightness, contrast, focal depth) are personalized not only to users' preferences but also to contextual task demands, drawing on real-time sensor data and AI-driven user modelling.

Perceptual Feedback and Immersion

Beyond ergonomics, perceptual feedback mechanisms play a critical role in enabling action–perception coupling in XR environments. These mechanisms bridge the gap between users' intentions, physical movements, and the system's responses, enhancing realism and embodiment. The SUN platform employs a multimodal approach, combining visual, auditory, haptic, and proprioceptive feedback. For instance, in rehabilitation scenarios, users receive visual feedback on their joint alignment or trajectory through avatars, reinforced by vibrotactile cues via wearable devices. Such multisensory integration has been shown to increase user engagement, reduce cognitive load, and improve motor learning outcomes in therapeutic settings [SUN D7.2 2024]. In the SUN pilot for lower-limb rehabilitation, for example, patients synchronize their movements with a visual avatar while receiving real-time performance corrections through both gaze-contingent overlays and tactile feedback. This layered feedback loop allows the system to accommodate sensory deficits, increase motivation, and facilitate neuroplastic responses essential to recovery [SUN D3.1 2024].

The Role of Gaze and Eye Tracking

A critical modality in perceptual feedback is gaze. Eye tracking allows for adaptive interface design (e.g., foveated rendering), intuitive interaction (e.g., gaze-based selection), and cognitive state inference (e.g., attention and fatigue estimation). Within SUN, gaze data is used not only to control interaction but also to infer perceptual load and adjust task difficulty accordingly [SUN D4.1 2024]. For instance, in complex industrial training scenarios, prolonged fixations or erratic saccades may indicate confusion, triggering adaptive help prompts or visual simplification. SUN's inclusion of gaze-based interaction highlights the project's commitment to reducing cognitive and physical barriers in XR environments and ensuring that perceptual demands are aligned with individual capabilities and context.

Toward Empathic and Inclusive Interfaces

Ultimately, perceptual feedback in SUN is not a purely technical function; it is a medium of empathy and inclusion. By modelling how users perceive and act within virtual spaces, SUN enables systems that are not only efficient but also sensitive to the diverse

sensory and cognitive profiles of users. This is particularly critical in scenarios involving people with severe disabilities or apathy-related impairments, where traditional input/output methods fall short. Here, visual and multisensory feedback becomes a core vehicle for motivation, agency, and engagement. By grounding perceptual feedback in principles of visual ergonomics, neurophysiology, and human-centered interaction, SUN advances the state of the art in immersive technologies, shaping environments that are not just immersive but truly human-aware.

19.3.2 Implications for Safety, Empathy, and Learning

Safety, empathy, and learning are deeply interlinked in XR environments, especially when deployed in healthcare, workplace safety, or assistive contexts. Within SUN, these aspects were addressed in an integrated way, with consortium-wide discussions ensuring that scenarios were designed from the outset to be convenient and intuitive to navigate.

The SUN project adopted a holistic understanding of safety that extended beyond physical risk management to include emotional reassurance and empathetic system design. In rehabilitation and assistive contexts, user confidence and perceived emotional safety are crucial enablers of learning and engagement. By creating safe environments that help to mitigate anxiety, XR reduces the cognitive load of unfamiliar interactions and fosters a sense of agency. Applications should not only promote motor learning and task completion but also reinforce trust and motivation. This highlights the importance of designing XR scenarios where safety is both a technical and an emotional construct. From a therapeutic perspective, safety and empathy are essential for effective learning and rehabilitation. Empathetic design enhances confidence, accelerates motor learning, and increases adherence to rehabilitation tasks. It can manifest through gentle feedback, supportive avatars, and transparent system intentions. When safety is embodied with empathy, XR experiences not only become safer but also more engaging and effective.

19.4 Ethics, Acceptability, and Human Values in XR Scenarios definition

19.4.1 Key Ethical Challenges in SUN scenario: Autonomy, Inclusion, Trust

The design and deployment of XR technologies, particularly when integrated with Artificial Intelligence (AI), must be grounded in strong ethical foundations—especially when these systems are used in sensitive domains such as rehabilitation, workplace safety, and disability support. The SUN project recognizes that technical innovation must be coupled with human values to ensure meaningful, equitable, and responsible outcomes. As such, two interrelated ethical challenges emerge across the SUN scenarios: autonomy and inclusion.

Autonomy - Enabling Control Without Overdependence: Preserving user autonomy is central to the ethical deployment of XR systems. In rehabilitation scenarios, users must feel empowered to take charge of their progress, rather than becoming overly dependent on automated feedback or AI-driven recommendations. SUN addresses this by designing systems that act as augmentative, supporting decision-making without replacing the user's or therapist's judgment [SUN D7.3 2023]. For example, in the rehabilitation of individuals with motor impairments, SUN's wearable and gaze-based systems collect real-time performance data and provide corrective feedback. However, they are designed to avoid intrusive nudging or paternalistic automation. The system proposes but never imposes. Autonomy is further preserved by enabling users to adjust parameters (e.g., intensity of feedback, avatar behaviour, interface pacing) to match their comfort level and consent preferences. In workplace scenarios, autonomy relates to workers' agency in controlling how and when XR assistance is activated, e.g., while detecting improper PPE usage or navigating hazardous environments. Crucially, SUN ensures that automation does not override human decision-making or promote a surveillance dynamic that could be ethically questionable.

Inclusion: Designing for Diversity and Accessibility: Inclusion within XR systems is not limited to accessibility for people with disabilities—it also concerns cultural, linguistic, age-related, and socio-economic differences that affect how individuals interact with digital environments. SUN adopts a human-centred co-design methodology to ensure that its XR applications are inclusive by design, not retrofitted. In the third SUN pilot, for instance, XR is used to enhance communication and motivation in users with severe motor or verbal disabilities. This requires not only technical solutions, such as body-machine interfaces and haptic feedback, but also ethical sensitivity to the dignity,

emotional variability, and cognitive diversity of users. Inclusion here means creating systems that are respectful of individual differences, adaptive to unique contexts, and free from biased assumptions about ability or intent. Moreover, SUN considers digital accessibility standards (e.g., multimodal interaction, clear visual hierarchies, customizable UIs) while also acknowledging the ethical duty to avoid reinforcing digital divides, e.g., between high-end XR users and those with limited access to infrastructure or training.

19.4.2 Acceptability Frameworks and Societal Readiness

For acceptability frameworks regarding SUN, a patient-centred lens is essential in addressing readiness, trust, and societal fit. Acceptability should not be treated as a checkbox but should emerge through continuous, qualitative engagement. This means asking questions such as: Does this feel comfortable? Do users understand why it matters? Do they feel safe and protected? Could this be applied outside the clinic? When such human values are respected, societal readiness increases.

Through questionnaires, co-design workshops, and real-world testing, the project explored how users perceive system purpose, trustworthiness, and relevance. Concerns relating to accessibility, fatigue, data handling, and onboarding for older or digitally inexperienced users were central to this process. These insights were instrumental in shaping interface design, personalization settings, and informed consent protocols. Acceptability, therefore, emerges as a multi-dimensional measure of societal readiness that combines ethical alignment, emotional comfort, and contextual suitability.

Trust consistently emerged as the most critical factor influencing readiness. Here, Outdoor Against Cancer (OAC)'s role as a patient-focused Non-Governmental Organization (NGO) added tangible value: its long-standing credibility with cancer survivors and patient communities strengthened the perception that SUN's XR developments were genuinely guided by end-user interests. By accompanying participants during Pilot 1 and acting as a mediator between patients, clinicians, and developers, patients' associations provided a direct feedback channel. This made engagement safer and more meaningful for participants, while also improving the flow of honest feedback to the consortium. In the SUN consortium, OAC's longstanding track record in working with diverse patient populations meant that its involvement was not just symbolic. It reassured stakeholders that the project was grounded in real-world health advocacy experience, and that data would not be treated as a commercial asset but as sensitive personal information requiring explicit consent and ethical handling. From the societal readiness perspective, the presence of a trusted NGO partner signalled that patient needs and rights were integrated into the project governance from the very start, and

helps mediate technical innovation and public acceptance. Ultimately, patients' involvement underlines the practical and ethical value of integrating organisations that closely work with vulnerable populations into technology consortia. By helping bridge innovation with everyday care, and by ensuring that discussions of usability, safety, inclusivity, and privacy were rooted in real-world patient concerns, SUN was able to validate its XR applications far beyond clinical or lab settings. This strengthened both the ethical grounding and the societal readiness of the platform.

19.4.3 Data Protection, Privacy, and the Application of AI

The technical work and piloting in SUN raise legal and ethical concerns due to the inclusion of human participants, the processing of personal data, the application of AI systems, and the safety of medical devices. The relevant framework applicable to SUN is presented in the following subsections, each addressing a distinct thematic area.

Human participants in research: Patient autonomy, informed consent, and the right to refuse participation are fundamental. For participants unable to give full consent, additional measures following the Helsinki Declaration are required, such as obtaining approval from an independent, qualified individual.

Data protection in research and exploitation: XR devices, equipped with sensors, 3D cameras, and control devices, facilitate data exchange and gather large volumes of personal information to assess user conditions and surroundings. They collect usernames, account information, activity logs, purchase history, user interactions, preferences, and demographic details (e.g., age, gender, birthdate). Location data and video recordings provide insights into user behaviour, enabling user profiling. These practices raise concerns about unintended data use and implicit consent through mere device activation. The technical capabilities and broad adoption of XR devices create unique privacy risks, including biometric data integration, haptic responses, real-time data gathering, and bystander exposure. Devices may also capture emotional states, moods, or personality traits to evaluate health and well-being. Principles of data minimisation, security, accountability, and informed consent are essential to address these risks.

Wearable devices and privacy: In SUN pilots, wearable devices monitor health conditions (e.g., blood pressure, heart rate), daily routines (e.g., XR for work), and well-being. These data streams support personalized movement guidance and lifestyle monitoring. However, the volume and sensitivity of collected information can intrude on private lives, including that of bystanders, as devices operate in both public and private spaces (homes, hospitals, shelters, places of worship). Robust safeguards, anonymisation, and transparent communication must ensure secure and ethical use of such data.

Artificial intelligence in SUN: AI techniques are used to process XR data, including raw scans, sparse 3D points, and incomplete acquisitions, enabling the generation of high-quality 3D meshes. AI supports the exploration of unknown environments in resource-constrained scenarios, and neural generative models enhance image resolution and complete 3D scene renderings by extrapolating context-based information. XR systems also employ algorithms for controller- and gesture-based controls, leveraging spatial data from head-mounted displays (HMDs). These functions enhance interactivity but expand the scope of data processing. The main concern is the "black box" nature of deep learning, which limits transparency and predictability, undermining human control and trust. According to the EU Guidelines on Trustworthy AI (2019), AI must be:

- *Lawful*, complying with all applicable laws and regulations;
- *Ethical*, ensuring adherence to ethical principles and values;
- *Robust*, technically and socially reliable.

Four ethical principles are central: respect for human autonomy, prevention of harm, fairness, and explicability. These requirements go beyond formal compliance, ensuring AI systems empower rather than manipulate, remain safe and secure, and communicate their purpose and decisions transparently.

19.4.4 Expected Lessons from SUN Scenarios and Ethical Governance

Ethical governance in XR and AI-based healthcare technologies must address the same core principles that guide patient advocacy: transparency, fairness, autonomy, and human connection. One of the central takeaways from the SUN project is that governance cannot remain an abstract policy layer but must be lived throughout the design, testing, and deployment of XR systems. Through participatory contributions from a multi-perspective consortium, SUN surfaced practical governance lessons grounded in scenario execution. These included the importance of proportional consent processes, user control over feedback intensity, and clear onboarding for XR interactions involving biometric data or AI-driven feedback. Beyond technical safeguards, the role of emotional and cognitive safety emerged as a governance priority in its own right, especially in rehabilitation and disability contexts. These experiences demonstrated that user trust and ethical accountability are mutually reinforcing elements of sustainable innovation.

At the HELT Symposium 2025, where the SUN consortium presented patient-focused perspectives on AI in healthcare, key lessons were distilled that resonate directly with the project's ethical approach:

- Patient autonomy over personal data must be safeguarded through privacy-by-design and granular consent mechanisms;
- Bias mitigation is essential to avoid unequal access or misdiagnosis across demographic groups;
- Corporate influence over healthcare AI must be balanced with public interest governance, avoiding "techno-feudal" dynamics where few entities control infrastructure;
- Preserving human connection is critical; technology should augment, not replace, therapeutic relationships;
- Open-source and collective frameworks can enhance transparency, public scrutiny, and trust.

For SUN, these principles meant that governance structures had to ensure patient representation in decision-making, mandate transparent development practices, and maintain ethical oversight throughout deployment. The interdisciplinary team's role was pivotal in anchoring these lessons in real-world advocacy experience, reminding the consortium that patient rights and dignity are not negotiable but foundational.

Looking ahead, the SUN experience demonstrates that embedding governance early, through, for example, patient advocacy, stakeholder diversity, and ongoing ethical reflection, can create XR systems that are not only functional but also trusted and socially acceptable. This precedent sets a strong foundation for future European projects to make ethical governance a standard practice, rather than merely an afterthought.

19.5 Evaluation, Impact Pathways, and Lessons Learned

19.5.1 KPIs, Metrics, and Data Collection from Scenarios

The evaluation strategy within the SUN project is grounded in a comprehensive framework of Key Performance Indicators (KPIs), metrics, and multimodal data collection protocols. These are not only technical or quantitative benchmarks but are also tightly coupled with human-centric values such as usability, inclusiveness, trust, and well-being. Given the project's focus on three interrelated pilot domains (rehabilitation, workplace safety, and interaction for individuals with severe disabilities), KPIs are context-sensitive, aligned with both scenario-specific goals and broader societal impact pathways.

Defining KPIs for Human-Centered XR: In rehabilitation scenarios (Pilot 1), KPIs include improvement in patient motor control, adherence to treatment protocols, and self-reported usability of XR interfaces. Quantitative metrics such as joint range of motion, time-to-completion of exercises, or EMG signal quality are paired with subjective reports via standardized tools such as the DASH (Disabilities of the Arm, Shoulder, and Hand) and WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) questionnaires.

In the industrial safety context (Pilot 2), KPIs measure not only system performance, e.g., accuracy of PPE detection or latency in obstacle recognition, but also human factors such as perceived safety, reduction of near-misses, and compliance with safety procedures. Cognitive workload, distraction levels, and trust in alerts are evaluated through in-situ observations and pre/post-task assessments.

Pilot 3 focuses on users with severe motor or communication impairments. KPIs include levels of engagement, successful communication interactions via avatars, and improvement in self-reported affect or motivation. In cases involving apathy, affective state recognition (via facial analysis or physiological sensors) is tracked alongside interaction frequency and duration.

Phased Evaluation Strategy: SUN follows a two-phase evaluation model: the first phase involves controlled experimentation with healthy volunteers or internal staff to test technical feasibility, system stability, and refine user interface elements. The second phase focuses on real-world deployment with target users (e.g., patients, factory workers, individuals with disabilities), collecting data longitudinally to measure impact and validate improvements over time[SUN D6.1 2024]. This phased approach allows iterative refinement and ensures that KPIs are grounded in authentic, user-driven experiences rather than abstract technological assumptions.

KPI Interpretation and Adaptive Feedback: A distinguishing element of the SUN evaluation strategy is that KPI tracking is not merely summative but also formative; data is used to adapt system behaviour in real-time. For instance, if a user demonstrates signs of fatigue, frustration, or poor motor control, the XR task may dynamically reduce difficulty, pause, or provide more supportive feedback. Similarly, in industrial settings, repeated PPE non-compliance may trigger a different kind of prompt or social interaction to promote safety culture. Such adaptive interpretation of performance metrics ensures that KPIs do not become static targets but instead function as part of a continuous improvement loop, both for the system and for its human users.

19.5.2 Societal Impact and Stakeholder Feedback

The societal impact of SUN extends beyond its technical achievements to encompass how its technologies were received, interpreted, and valued by diverse stakeholders. Structured stakeholder engagement was facilitated across all pilot domains, with all the team supporting patient involvement in the rehabilitation use case. Feedback gathered from patients, clinicians, caregivers, and therapists highlighted the motivational potential of XR environments, the perceived safety and ease of use, and the possibility of integrating these tools into daily care routines. Importantly, participants expressed trust in the system due to its adaptability and the presence of human accompaniment. These insights confirmed that SUN's approach, which is grounded in co-creation and ethical design, can deliver XR applications that achieve real-world legitimacy and social relevance. XR tools built with human values extend beyond clinical utility into broader societal impact, while societal readiness improves when end-user perspectives are embedded directly into design and governance.

A general networking aspect and improved informational exchange between various stakeholders further enhanced the project's societal relevance. The richness of perspectives, backgrounds, and competencies fused and interlinked within this endeavour was key to its success. The exchanges within the SUN consortium, complemented by the External Stakeholder Advisory Board (ESAB), shaped and influenced the definition and assessment of the Key Innovation Results (KIR). From the perspective of survivors, issues such as usability, emotional engagement, safety, human accompaniment, inclusivity, data and privacy concerns, accessibility for cancer survivors with fatigue, and tailored onboarding for older or digitally inexperienced users were identified as crucial. Addressing these concerns helped validate XR applications in contexts beyond clinical or lab settings. This directly supported SUN's mission to deliver not just functional XR technologies, but meaningful and equitable solutions that benefit humans first.

While a full assessment of societal impact will only be possible after project completion, the outlook is very promising. By embedding patient advocacy, ethical design, and stakeholder dialogue from the outset, SUN has laid the groundwork for the adoption of its platform and technologies on larger scales to strengthen both trust and societal readiness for immersive technologies.

19.5.3 Barriers, Enablers, and Adaptive Methodologies

The development and validation of human-centred XR systems—especially when integrating AI, rehabilitation protocols, industrial safety procedures, and accessibility tools—requires a nuanced understanding of both systemic barriers and contextual enablers. In the SUN project, these challenges are not treated as incidental but as design drivers.

To this end, SUN adopts a dynamic and reflexive methodology that adapts to the evolving needs of diverse users, stakeholders, and settings.

Barriers to Human-centred XR Implementation: Several key barriers have emerged across the three SUN pilot domains:

- *Technological Complexity and Fragmentation*: XR systems involve multiple hardware and software components, wearables, sensors, AI models, and rendering pipelines. This complexity creates integration risks and steep learning curves for end users, especially therapists, factory operators, or non-technical caregivers;
- *Accessibility and Physical Constraints*: In Pilot 3, addressing users with severe motor or verbal impairments, many existing XR tools were found insufficiently accessible or physically cumbersome. Traditional input methods (e.g., hand gestures, button clicks) often exclude users with limited mobility. Hardware ergonomics and comfort also remain a significant constraint;
- *Trust and Ethical Sensitivities*: In both healthcare and workplace scenarios, there is hesitancy toward AI-driven automation and data collection. Workers may perceive PPE tracking or safety alerts as surveillance. Patients may worry about depersonalized care or privacy violations when using body-machine interfaces;
- *Resistance to Change*: In both industrial and healthcare contexts, the introduction of immersive technology may face inertia due to organizational culture, lack of digital readiness, or competing priorities. For example, therapists may prefer conventional protocols; factory staff may view XR as disruptive.

Despite these challenges, the SUN project has identified several enablers that support the successful deployment and adoption of XR-AI systems:

- *Participatory Co-Design*: Involving therapists, patients, industrial workers, and users with disabilities early in the design process ensures solutions reflect real-world needs and constraints. This process builds trust, relevance, and user acceptance;
- *Ethical Governance and Transparency*: Ethical impact assessments, informed consent protocols, and privacy-by-design strategies are embedded throughout the project. This mitigates risk and enhances societal readiness, especially where vulnerable groups are involved;
- *Adaptive Personalization*: SUN leverages AI to dynamically adjust the complexity, intensity, and sensory modalities of tasks in real time. For example, a rehabilitation task can be made easier if the system detects fatigue or stress, while still recording clinical progress;

- *Training and Capacity Building*: Workshops, hands-on testing, and dedicated support sessions with practitioners (e.g., occupational therapists, line supervisors) have proven crucial in lowering barriers to adoption. By investing in user competence, the project fosters long-term sustainability of the XR ecosystem;
- *Adaptive Methodologies in SUN*: Rather than assuming fixed experimental conditions or user behaviours, SUN iteratively refines its tools and protocols in response to ongoing feedback and real-world data;
- *Hybrid Evaluation Models*: SUN combines qualitative (interviews, focus groups, observational notes) and quantitative (biometric data, task performance, self-report questionnaires) approaches. This mixed-methods strategy ensures technical KPIs are matched with experiential insights;
- *Contextual Adaptation*: Recognizing that XR usage is always situated, whether in a hospital room, factory floor, or domestic setting, SUN allows for flexible configurations and data collection strategies depending on space, hardware, or user needs;
- *Reflexivity and Feedback Loops*: Feedback is not only collected for end-of-project evaluation but continuously used to reconfigure tasks, update avatars, fine-tune haptic responses, or revise user interface elements. This promotes alignment between system evolution and user reality.

19.6 Future Perspectives

The outcomes of the SUN project open up multiple directions for sustainability, accessibility, and broader societal adoption of XR. The following perspectives highlight how lessons learned from pilots can evolve into long-term strategies, shaping both practice and policy in the years to come.

Sustainability as a social fabric: Sustainability in XR cannot be reduced to software updates or hardware maintenance cycles. It must be seen as a living fabric woven across patients, workers, caregivers, and organizations. The SUN pilots showed how communities themselves can act as sustainability anchors. Peer-led onboarding, long-term training, and feedback-driven adjustments ensured that technologies did not remain static prototypes but evolved as responsive tools aligned with users' changing needs. A key future direction is to embed patient associations, worker unions, and disability advocacy groups directly into governance structures, so that they not only benefit

from the outcomes but also actively shape their evolution. In this participatory vision, sustainability is safeguarded by trust, empowerment, and shared ownership.

Toward accessibility and community-based adoption: For human-centred XR to scale, barriers of cost, usability, and physical accessibility must be decisively lowered. Future deployments must guarantee that hardware ergonomics and interaction modes do not exclude those with limited mobility or sensory impairments. Establishing local adoption hubs—in hospitals, NGOs, rehabilitation centres, or community facilities—represents a practical step forward. These hubs could provide access to equipment, structured training, and shared peer-to-peer knowledge exchange. Beyond increasing adoption rates, they would extend the societal footprint of XR, making it a familiar and supportive presence in daily life rather than a niche or experimental tool.

Scaling Pilots into replicable models: The three SUN pilots proved the feasibility of XR-AI systems in highly contextualized environments. Importantly, each pilot produced transferable assets: modular XR components, interoperable data pipelines, user training protocols, and domain-specific KPIs. These can now be recombined and tailored for new applications, from home-based rehabilitation for chronic conditions to safety monitoring in logistics and construction, or digital inclusion in education and communication for people with cognitive or mobility impairments. The project's modular architecture supports this scalability, allowing gesture recognition, gaze tracking, haptic feedback, and AI-driven user modelling to be flexibly adapted across sectors and use cases.

Cross-sectoral synergies and innovation pathway: By integrating XR with AI, ethics, and participatory design, SUN lays the groundwork for cross-sectoral innovation. In healthcare, immersive technologies can extend into telemedicine, remote diagnostics, and patient education, enhancing personalization and equity in care delivery. In education and training, embodied XR experiences can offer gamified, adaptive environments for vocational re-skilling and cognitive support. In Industry 5.0, XR combined with IoT and human-in-the-loop AI promises safer, more productive hybrid workplaces where human well-being is prioritized alongside efficiency. These synergies illustrate that XR is not a closed niche technology, but a platform for systemic transformation across domains.

Policy, Design, and Research Directions: Future perspectives also point to the urgent need for coordinated action at policy and design levels. XR introduces unique challenges around embodiment, biometric capture, and immersive consent that extend beyond current AI legislation. Policymakers must therefore craft XR-specific ethical and legal frameworks that recognize immersive data—such as gaze paths, gesture profiles, and spatial behaviours—as sensitive and worthy of sovereignty protections. Designers, meanwhile, must embed multimodal accessibility, participatory co-design, and trans-

parent consent mechanisms into the heart of system development. Researchers should continue to explore adaptive and reflexive evaluation methods that balance technical KPIs with experiential insights on trust, autonomy, and inclusion. In this sense, the lasting legacy of SUN may not be any single technological breakthrough, but its model of responsible, participatory, and ethically resilient innovation—a blueprint for how Europe can lead in building XR systems that are not only technologically advanced but also deeply human.

19.7 Conclusions

The SUN project demonstrates that XR and AI, when guided by ethical intelligence and human-centred values, can move beyond experimental showcases and become trusted instruments of societal transformation. However, ensuring that such systems evolve sustainably requires more than technical refinement: it demands a long-term commitment to inclusivity, accessibility, and governance grounded in human dignity. The future of human-centred XR will hinge on how effectively technology can be integrated into the rhythms of everyday life while respecting autonomy and trust. SUN has shown that XR is not just about immersion, but about empowerment—helping patients recover mobility, workers operate more safely, and people with disabilities participate more fully in society. The next challenge lies in ensuring that these advances do not remain isolated pilots, but scale into sustainable ecosystems, co-owned by communities and governed by ethical foresight.

References

Leonardis, Daniele and Serra, Federica and Camardella, Cristian and Sarri, Froso and Kasnesis, Panagiotis and Toumanidis, Lazaros and Contiero, Amalia and Mendez, Vincent and Muheim, Jonhatan and Symeonidis, Spyridon and Diplaris, Sotiris and Xefteris, Vasileios-Rafail and Poulios, Ilias and Vrachnos, Panagiotis and Loupas, Georgios and Sarakatsanos, Orestis and Paraskevopoulos, Ioannis and Inguglia, Alessandro and Vitucci, Giuseppe and Carrara, Fabio (2024). *D3.1 - Human Machine Interaction Components Specifications*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

Perossini, Fabio and Caracciolo, Giuseppe et al. (2024). *D7.2 - Impact pathway and dissemination V1*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

Posteraro, Federico and Bertolucci, Federica and Pietrini, Alice and Perossini, Fabio and Leonardis, Daniele and Amato, Giuseppe and Vadicamo, Lucia and Yao, Cong and Pamminger, Carina and Minich, Tom and Sevgili, Baris Ege and Rocamora, Pablo and Giménez, Alfredo and Mendez, Vincent and Léger, Bertrand and Kasnesis, Panagiotis and Georgoudis, George and Plavoukou, Dora and Paraskevopoulos, Ioannis and Vitucci, Giuseppe and Inguglia, Alessandro and Mula, Josefa and Pérez-Bernabeu, Elena and Martín, Xabier (2024). *D6.1 - SUN Pilot planning*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

Stan, Alexandru, George Ioannidis, Andrea Petrus, Fabio Perossini, Federico Posteraro, Alfredo Gimenez Millan, Bertrand Leger, Daniele Leonardis, Josefa Mula, Vincent Mendez, Panagiotis Kasnesis, George Georgoudis, and Ioannis Paraskevopoulos (2024). *Deliverable D2.2 - User requirements and scenarios*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

Symeonidis, Spyridon and Diplaris, Sotiris and Vrachnos, Panagiotis and Loupas, Georgios and Xefteris, Vasileios-Rafail and Callieri, Marco and Di Benedetto, Marco and Giorgi, Daniela and Palma, Gianpaolo and Leonardis, Daniele and Kasnesis, Panagiotis and Joyce, Leesa and Mania, Katerina and Sarri, Froso and Paraskevopoulos, Ioannis and Inguglia, Alessandro and Vitucci, Giuseppe (2024). *D4.1 - 3D Acquisition and Real-time XR Visualization Components Specification*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

Yao, Cong and Paul Quinn (2023). *D7.3 - Ethics and the GDPR*. Tech. rep. SUN project - Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR). URL: https://www.sun-xr-project.eu/sun-public-deliverables/.

20. Extended Reality for Rehabilitation

Federica Bertolucci[1], and Federico Posteraro[1], Panagiotis Kasnesis[2], Amalia Contiero Syropoulou[2], Lazaros Toumanidis[2], Alessandro Inguglia[2], Ioannis Paraskevopoulos[2], Theodora Plavoukou[2], Giorgos Georgoudis[2], Vasileios-Rafail Xefteris[3], Spyridon Symeonidis[3], Sotiris Diplaris[3], Stefanos Vrochidis[3],Froso Sarri[4], George Ramiotis[4], Michael Effraimidis[4], Katerina Mania[4], Daniele Leonardis[5] ,Cristian Camardella[5] , Federica Serra[5]

[1] Azienda USL Toscana Nord-Ovest, Italy
[2] ThinGenious, Greece
[3] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece
[4] Technical University of Crete
[5] Scuola Superiore Sant'Anna (SSSA), Italy

Abstract. This chapter explores the use of Extended Reality (XR) in the rehabilitation of orthopedic, neurological, and oncological conditions, with a focus on the development and validation of a novel XR-based platform designed to enhance patient engagement and recovery outcomes. The platform integrates motion sensors, electromyographic sensors, haptic actuators, and real-time feedback mechanisms to support both supervised and personalized rehabilitation. We present two key clinical use cases: 1) Upper Limb Rehabilitation, focusing on shoulder pathologies and lymphedema post-breast cancer treatment, and 2) Lower Limb Rehabilitation, targeting knee recovery. Both scenarios leverage XR technologies, such as augmented reality headsets and 3D avatars, to create interactive, goal-oriented exercises. The system's ability to provide real-time, performance-related feedback to patients and clinicians ensures personalized care while maintaining the therapeutic relationship between the patient and therapist. Pilot studies conducted with healthy volunteers and patients revealed high levels of usability, patient satisfaction, and promising clinical outcomes, including improved upper limb function and reduced lymphedema. Although the results indicate the feasibility of XR integration into rehabilitation, further refinements are needed, particularly in enhancing

avatar accuracy and AI-based feedback. This chapter highlights the potential of XR technologies to revolutionize rehabilitation practices by increasing patient motivation, supporting personalized care, and enhancing therapeutic outcomes in both clinical and home settings.

20.1 Introduction

The key focus of this chapter is to provide innovative solutions for the rehabilitation of orthopaedic, neurological, or oncological conditions. These pages detail the work dedicated to make rehabilitation sessions appealing, engaging, and stimulating from a social, playful, and cognitive perspective, maintaining direct patient-therapist communication. The chapter also describes the clinical advantages offered by the integrated system composed of motion sensors, electromyographic sensors, and haptic actuators providing feedback both to clinicians and to the patients on the correctness of the exercise execution.

20.1.1 The Challenge: Motivating Patients and Personalising Rehabilitation

Functional recovery following various orthopaedic, neurological, or oncological problems can take a very long time, requiring demanding, repetitive, and prolonged rehabilitation interventions that make it difficult for patients to adhere to the treatment plan adequately. Furthermore, it is not always possible to access personalized treatments that take into account individual patient progress, resulting in greater effectiveness with fewer resources

20.1.2 The Solution: An XR Integrated Platform for Supervised Personalised Rehabilitation

To address these challenges, we have developed an XR platform to motivate patients to exercise efficiently by providing them and the physiotherapist with feedback on their performance. This is accomplished by exploiting the developed machine learning-based SUN modules to perceive rehabilitation-related metrics (e.g., number of repetitions, range of movement) and by providing feedback through the movement of their avatar and the interaction with 3D virtual objects.

20.1.3 Key Use Cases

The platform's capabilities were applied in two clinical use cases, each targeting a specific aspect of the overall challenge.

Use Case 1: Upper Limb Rehabilitation scenario

The Upper Limb Rehabilitation scenario focuses on recovery of upper limb function after shoulder pathologies/injuries or Lymphedema due to breast cancer.

In this use case, the system is designed to be used in a clinical setting where the patient and therapist are co-located. The patient sits at a desk, which serves as the physical anchor for the virtual workspace presented through an Augmented Reality (AR) headset. The AR environment allows the patient to interact with virtual objects overlaid on the real environment, allowing the therapist to remain visible and actively involved during the session.

The virtual rehabilitation tasks are embedded within a serious game framework, where the patient performs goal-oriented exercises. These are implemented in the form of a pick-and-place task, moving virtual objects onto bookshelves, progressively increasing in height and lateral displacement.

The AR environment displays the interaction elements implementing these tasks, providing tuning of key parameters of the exercises, i.e., the height and width of the workspace and target object positions. Moreover, the patient has the ability to preview the exercise via an avatar demonstration before performing the protocol.

The system is also conceived to provide real-time feedback on the execution quality, such as identifying and warning the user about postural errors or compensatory movements. This is supported by the addition of a vision-based pose estimation system for tracking the shoulder, elbow, and torso leaning angles (described in Chapter 13). Also, an emotion recognition module, based on a wearable biometric sensor, is implemented to monitor the effort of the patient while performing the exercise (described in Chapter 15).

To enhance immersion and support motor learning, wearable haptic devices on the fingertips and wrist, respectively, are introduced. Fingertip units (described in Chapter 7) deliver tactile feedback that corresponds to virtual interactions, such as contact cues and pressure sensations that simulate contact with objects. Skin stretch cues at the wrist (described in Chapter 8) are used to convey information about the posture of the arm held during the exercises, in particular addressing the typical compensatory movement expected during the pick-and-place task. The two haptic methods have been combined and implemented in a single glove for this Case Study.

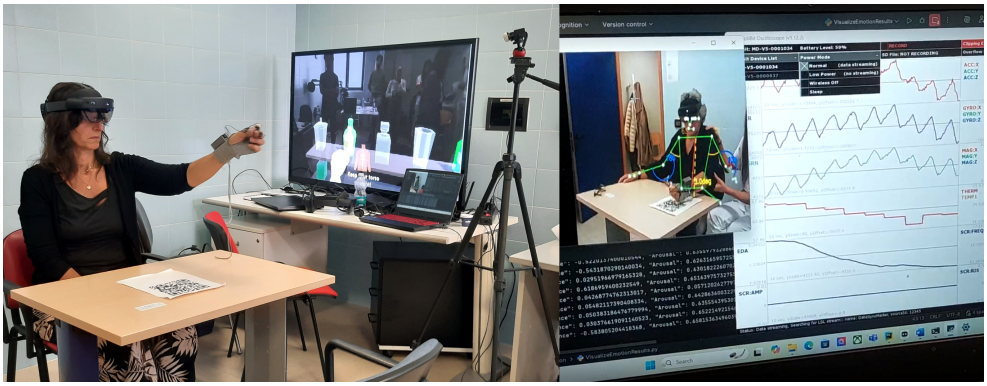


Figure 20.1: (Left) the clinical setup for the Use Case 1: Upper Limb Rehabilitation, including the XR visualization of the virtual exercise (streamed to the external monitor), the haptic glove for interaction and postural tactile hints, the Emotibit armband, and the external camera for upper limb tracking and biometric signals recording. (Right) screenshot of the real-time streaming of the measured data, showing upper limb tracking and physiological signals.

All these elements combine to create a flexible, responsive rehabilitation system that can be adapted based on the patient's ongoing performance, while preserving direct interaction and oversight from the therapist. An overview of the experimental setup for the Upper Limb Rehabilitation scenario is shown in Figure 20.1.

Use Case 2: Lower Limb Rehabilitation scenario

In the lower limb rehabilitation scenario, the system is designed to be used as a virtual supervisor of the patient, with the physiotherapist only observing the rehabilitation process. The patient is asked to perform three common knee rehabilitation exercises: a) squat, b) seated leg extension, and c) walking (for gait analysis).

The AR environment allows the patient to observe their movements that are displayed by a 3D avatar that mimics them. Moreover, the system also provides feedback to the user regarding whether they perform the exercise correctly and, if not, what went wrong (see also Chapter 12 and Chapter 13). Through this real-time feedback and the developed XR immersive environment, SUN aims to motivate the patients and also assist them in progressing the rehabilitation journey more efficiently. Figure 20.2 depicts the lower limb setup. A patient performs a seated leg extension exercise, while being able to see her avatar performing the same exercise (in mirror view). The green bar on top of the avatar provides information about the repetitions made and the correctness of each repetition.

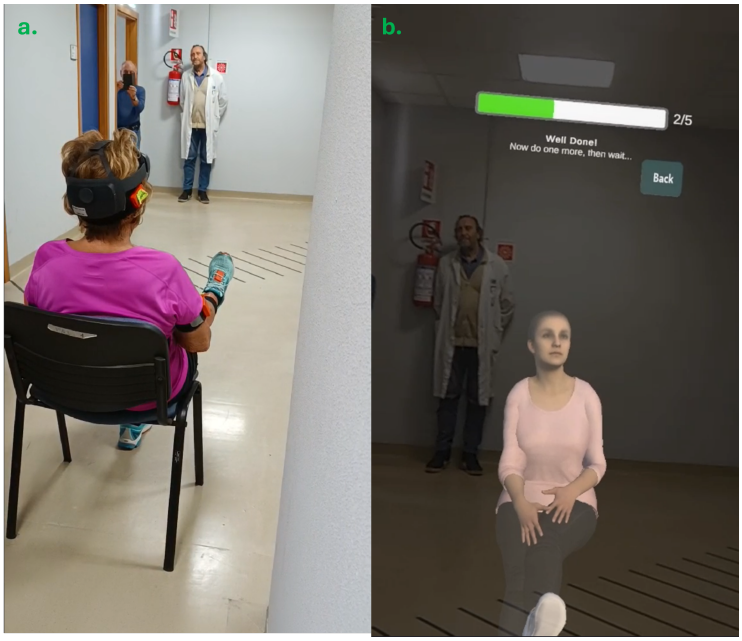


Figure 20.2: The lower limb setup shown in the (Left) and the patient's view displayed at the (Right).

The system uses a hyper-realistic, animatable 3D human avatar, generated through the pipeline developed within the SUN project and described in Chapter 2.

20.2 Methodology and Results

The development and evaluation of the SUN platform for this pilot follows a structured approach encompassing system design, component integration, and a formal validation phase in turn divided into two phases: an initial test on healthy volunteers and then the final test on patients.

20.2.1 System Architecture and Implementation

For the Upper Limb Case, a gamified scenario was developed in AR that aligns with the recovery of the targeted motor function. The user interacts with virtual 3D objects, picking them up and placing them in a vertical structure, resembling shelves, at incremental heights, starting from desktop level and reaching just above the head. On each

height, there are shapes indicating target positions where the user must correctly match and insert the corresponding 3D objects. As the user fills the shapes, on each height, they progress incrementally to the next height. The system provides color-coded visual hints that indicate errors regarding target placement (red) and object matching (orange). A dedicated desktop GUI application was developed to configure the scenario parameters, including the number of shelves (height), the number of objects that can be placed on the shelves, and the limb (left or right) that should perform the rehabilitation. The GUI app also displays data from external components.

Throughout application use, haptic feedback references are transmitted to the wearable devices based on virtual object interaction events and input from the computer vision-based pose estimation module. Emotion recognition data captured from the user is displayed in the desktop application. Additionally, the system provides the ability for therapists to join the same virtual environment as the patient, observe the patient's interactions, and interact directly utilizing dynamic cues to support object placement tasks (see also Chapter 9). Finally, interactions with the XR application's UI could be performed via hand and gaze input. For the gaze modality, users can confirm the selection ("click the button") with a blink to avoid misinteractions.

For the Lower Limb case, the AR environment centers around the avatar, through which the results of the ML components are displayed. Real-time pose estimation data are mapped to the avatar to mirror user movements. Augmented feedback is displayed on the avatar based on the postural assessment component, indicating the correctness of the performed movements. For the squat and seated knee flexion and extension exercises, the avatar is placed in a mirrored position in front of the patient, while during gait training, the avatar is positioned ahead of the patient, shown from a rear perspective to simulate walking in the same direction. Before each exercise, the avatar demonstrates the correct movement through predefined animations

The pilot architecture leverages a multi-layer architecture including four layers: a) data source layer, b) information layer, c) application layer, and d) presentation layer. Figure 20.3 presents the modules that were included in the 2nd validation phase. In terms of data sources, we exploited Inertial Measurement Unit (IMU) and surface Electromyography (sEMG) sensors for the postural assessment and the wearable-based pose estimation. Their results are sent to the XR application by passing through OmniBridge (see Chapter 16). The results of the pose estimation are depicted using the 3D avatar, while the whole application runs on the AR device. Moreover, camera input is used to estimate the angles of the user (i.e., elbow, shoulder), and a wristworn device (i.e., EmotiBit) is used to infer the user's emotion. This information is also sent to the XR application exploiting OmniBridge. Moreover, the XR app lets the user interact

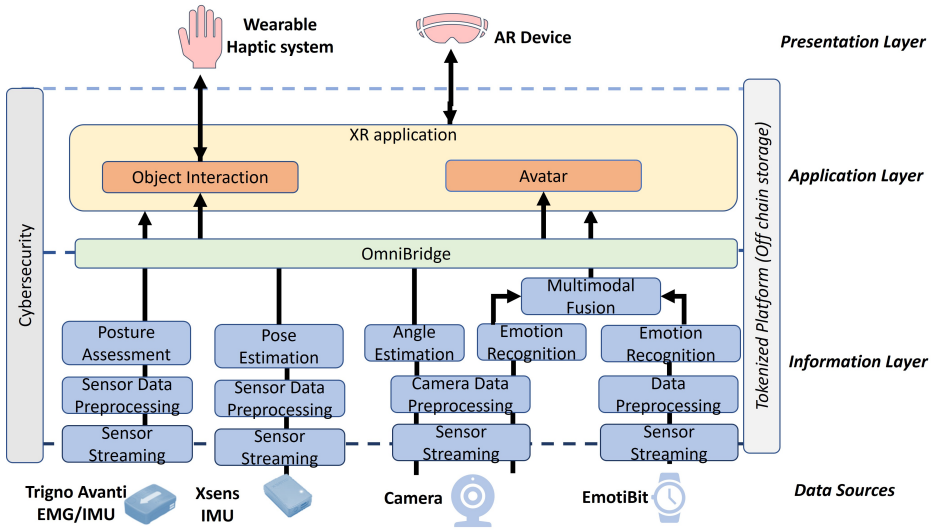


Figure 20.3: Abstract SUN platform architecture for Pilot 1.

with virtual objects and communicate with the developed wearable haptic system to increase the user's immersion levels.

20.2.2 Validation Experiment

The validation of the pilot took place at Versilia Hospital in Lido di Camaiore (Lucca, Italy) and consisted of two phases: a preliminary phase where the first release of the SUN XR platform was tested on healthy volunteers (Figure 20.4 and Figure 20.5), and a final validation phase where the final version of the platform was tested on real patients as an innovative rehabilitation tool (Figure 20.1 and Figure 20.2).

While the goal of the first validation phase was to test the integrated platform in terms of ease of use and acceptance by clinical staff and naive healthy subjects, the goal of the final validation was to test the new XR technologies on patients with orthopaedic/oncologic pathologies, not only in terms of acceptability but also in terms of rehabilitation clinical impact.

Experimental Tasks

Upper limb The user was seated in a comfortable position, with their arms resting on a table and their elbows bent at approximately a 90° angle. The user wore the haptic

glove and was given instructions on the actions to be performed. A demonstration of the exercise was available for the user to view before starting the exercise.

The exercise consisted of training the pick-and-place functionality. The subject had to reach targets in different areas and heights of the workspace. A game scenario was developed in an AR environment that aimed to train the upper limb motor function. The user was able to view the environment and perform the exercise through the AR HMD device HoloLens 2.

The AR application consisted of a game in which the subject must pick up virtual objects and place them in a vertical virtual structure. The scenario focuses on the repeated practice of reaching movements in different directions on the vertical plane. While the person is sitting, 3D objects are presented in front of them at the height of the desk. The subject must choose each of the objects and place them on specific targets on the vertical plane. Each 3D object corresponds to a particular target position, and as the subject manages to place each object, the level of the shelf gradually becomes higher. During the exercise, the subject is presented with the target position of the object to be placed.

The haptic glove worn by the user enables the rendering of touch and pressure sensations by means of an actuated band in contact with the fingertips. In addition, haptic cues are delivered at the wrist to inform the user about possible incorrect posture of the arm, addressing typical compensatory movements expected in similar pick-and-place tasks. In case the arm is not fully extended during the target reaching phase, skin stretch cues are delivered, with intensity proportional to the extension error, suggesting the user to extend the arm rather than lean the trunk toward the target object.

Furthermore, the user was equipped with a wearable device, the Emotibit, for measurement of biometric signals. The EmotiBit wearable system is a device equipped with photoplethysmography (PPG), heart rate, electrodermal activity (EDA), and temperature sensors. The device is open source and offers the flexibility of being worn almost anywhere on the body. For convenience, the users were asked to wear the EmotiBit on the wrist like a smartwatch. These sensors are able to detect and monitor the emotional state and the effort made by the users while performing the exercises, providing very useful information about how much the users are engaged in the exercises.

Lower limb The lower limb rehabilitation scenario consisted of three different exercises: sitting leg flexion and extension, squats, and gait training. In this scenario, the AR environment is focused on the avatar, a patient avatar that replicates the patient's movements and provides guidance. At the beginning of the session, one of the three exercises can be chosen through the application menu. For the squat exercise, the

Figure 20.4: A participant testing the Upper Limb Rehabilitation scenario during the first validation session at the Versilia Hospital.

avatar is positioned in a mirror image in front of the subject. Similarly, for the seated exercise, the avatar is positioned in front of the subject in a mirror image of the sitting position. For the gait training, the avatar is positioned in front of the subject, but is seen from behind, with a posterior perspective. Figure 20.5 shows a participant testing the lower limb application during the first validation phase.

For the walking exercise, a virtual corridor is displayed on the ground to delineate the physical space needed for training, while for the squat exercise, a virtual ring is displayed around the patient that delimits the space needed to safely perform squats. Before performing the exercise, the patient can observe the correct way to perform it through predefined animations shown by the avatar. During the exercise, the patient's movements are replicated by the avatar via the 3D pose estimation module. In addition, the system provides visual feedback on whether the patient has performed the exercise correctly based on data from the wearable-based posture assessment module.

Moreover, eight wearable sEMG sensors (Delsys Trigno Avanti), consisting of IMUs and sEMG inferred the correctness of the exercise performed and were interfaced with the virtual system to display the results.

Preliminary Validation Phase

Eleven healthy volunteers (consisting of a mix between the hospital department staff and other members of the SUN consortium) participated as end-users in the upper limb (9 subjects) and lower limb (2 subjects) scenarios, each performing one session

Figure 20.5: A participant performing a squat for the Lower Limb Rehabilitation scenario during the first validation session at the Versilia Hospital.

As assessment measures, kinesiological measures such as task execution time and joint range of motion (angle) were recorded for both upper and lower limb; sEMG signal was recorded for the lower limb.

Moreover, participants were administered an ad hoc questionnaire (the *Pilot 1 user questionnaire*) comprising several Likert-scale items (statements rated on a five-point scale ranging from 'strongly disagree' to 'strongly agree') addressing usability, performance, and acceptance by the users. The preliminary validation tests successfully demonstrated the stable integration of the core components.

Overall, participants reported a good acceptability of the system and satisfaction, found the sessions very engaging, and would recommend the use of the system to others.

Nevertheless, regarding upper limb sessions, some users complained that the glove did not fit perfectly for them (it was too large), and others reported that the release of the object was difficult. As for lower limb sessions, users found the avatar feedback not always accurate and therefore not having much impact on the performance at the moment. Analysis of the user feedback from the first validation phase provided key insights into the system's performance and areas for refinement in view of the clinical validation phase.

Final Validation Phase: Clinical Data from Rehabilitation Treatment

The first experimental validation paved the way for the clinical trial on real patients affected by orthopaedic or oncological diseases.

Therefore, for the phase of clinical validation, six patients were recruited at Versilia Hospital: two patients with upper limb lymphedema secondary to oncologic surgery, one patient with an orthopedic shoulder condition, and three patients who had recently undergone knee surgery. Each of these patients had already participated in conventional rehabilitation sessions according to standard clinical practice for their specific condition. Each of the recruited patients carried out one or more rehabilitation sessions using the devices of the SUN platform. Patients were assisted by the clinical staff of the Rehabilitation Unit of Versilia Hospital. As an assessment measure, specific validated clinical scales were administered to all patients. The study was approved by the local ethics committee, and all participants signed informed consent for participation in the clinical study and for data processing. Moreover, as in validation phase 1, in order to assess the XR integrated platform in terms of usability, performance, and acceptance by the users, the *Pilot 1 user questionnaire* has been administered to all participants.

Upper limb Rehabilitation Sessions Each of the three patients took part in a daily treatment session lasting approximately 40 minutes in total (including setup and treatment) for four consecutive days.

For the treatment, as in the first validation phase, the patients wore the HoloLens headset and performed grasping and releasing exercises with virtual objects placed on shelves of progressively increasing height. In accordance with the principle of XR, the sensation of grasping and releasing was enhanced by haptic feedback from the glove.

During the treatment, the patients were assisted by the physiotherapist (Figure 20.6), whose task was to correct any movement errors or incorrect postures. The physiotherapist could also wear another headset, allowing them to see in real time the same virtual objects displayed to the patient during the exercise.

Figure 20.6: A patient involved in the Upper Limb Rehabilitation scenario during the Final Validation Phase. The therapist was co-located in the same XR environment, enabling a natural interaction and effective physical assistance toward the goal-oriented motor task.

The outcome measures used were: the measurement of the upper limb circumference at different points (to assess the extent of lymphedema) and the DASH scale [DASH 2006; Beaton et al. 2001] to evaluate the functional limitation of the upper limb in performing activities of daily living. Outcome measures were recorded at the beginning and at the end of the last session.

Upper limb Rehabilitation - Clinical Results and Patients' Feedback Tests ran smoothly without any major technical issues, demonstrating the feasibility of the method. As a clinical result, arm circumference of patients affected by lymphedema slightly decreased, indicating a reduction of the edema, and the DASH score of the patients with shoulder impairment decreased, indicating a decreased limitation in daily activities. The results of the questionnaire showed high scores for almost every item, indicating a good performance of the system in the three areas of exploration: "User Experience & Usability", "Haptic Feedback & Glove Usability", "Overall Satisfaction & Future Use".

Patients reported an overall good experience with this innovative type of rehabilitative treatment. One of them reported that it was pleasant to receive the haptic feedback so that she knew she had grasped the bottle correctly. All patients reported a subjective feeling of improvement in upper limb function and motricity, but two of them complained that sometimes the "grasp and release" did not work properly and this disturbed their exercise. Regarding the postural feedback, one patient suggested

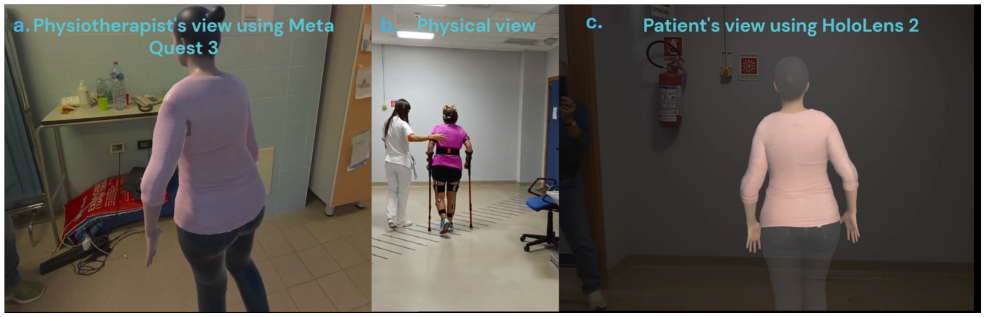


Figure 20.7: A patient involved in the Lower Limb Rehabilitation scenario during the Final Validation Phase. (b) One therapist assisted the patient in performing gait activity. (a) Another therapist was able to see the patient performing the exercise remotely. (c) The patient was also able to see her avatar performing the exercise in front of her.

adding more specific instructions about how to correct posture. One patient reported that it was a positive rehabilitative experience, but that it necessarily must be integrated with conventional physiotherapy.

Lower limb Rehabilitation Sessions For the treatment, as in the first validation phase, the patients wore the HoloLens headset and were asked to perform exercises of squatting (if possible), knee flexion and extension from a sitting position, and walking (Figure 20.7). sEMG activity of lower limb muscles and knee kinematics were recorded. The patients could see their avatar from behind performing the same exercises, and visual feedback indicating the quality of their performance. sEMG activity was monitored by clinicians during the exercises, while a report representing knee kinematics could be produced immediately after the session.

During the treatment, the patients were assisted by the physiotherapist, whose task was to correct any movement errors or incorrect postures (Figure 20.7b). The doctor could also wear another headset, allowing them to see in real time the avatar of the patient performing the exercises from a remote site (Figure 20.7a).

The outcome measures used were: Oxford knee score (OKS) [OKS 2016] and Western Ontario and McMaster Universities osteoarthritis index (WOMAC) [Dawson et al. 1998; Bellamy et al. 1988], evaluating the overall knee function and the patient's ability to manage normal activities.

Lower Limb Rehabilitation - Clinical Results and Patients' Feedback As for the upper limb, sessions ran without any major technical issues, demonstrating the feasibility of the method.

Pre-post results of clinical scales show a slight improvement except for WOMAC in one patient.

The results of the questionnaire showed high scores in the areas of "User Experience & Usability" and "Overall Satisfaction & Future Use", while, as for "Effectiveness of the Avatar & AI Feedback", one gave lower scores, indicating that this is an improvable aspect of the platform.

All patients stated that they had no difficulty using the XR system, including the older patients who were less familiar with technological devices. They found the experience interesting and engaging, and they willingly completed all the sessions. However, since they had also undergone conventional rehabilitation for their condition, they confidently stated that, at present, traditional physiotherapy cannot be replaced by virtual reality systems.

20.3 Conclusions

In the rehabilitation scenario, the integration of the various components outlined in the project (augmented reality, use of avatars, haptic feedback, and continuous monitoring systems of clinically relevant parameters) has been demonstrated in a real and sensitive environment such as a rehabilitation hospital. While the feasibility of the project was proven during the experimental validation phase, its actual applicability in the rehabilitation setting was confirmed during the clinical validation phase, with encouraging clinical results for future uses.

The clinical tests were in fact carried out without technical complications and with excellent compliance from the subjects, including the older participants who were less familiar with technological devices. The patients expressed a hope to be able to use such technologies in the future, if integrated with conventional physiotherapy.

References

Beaton, Dorcas E, Jeffrey N Katz, Anne H Fossel, James G Wright, Valerie Tarasuk, and Claire Bombardier (2001). "Measuring the whole or the parts?: validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity". In: *Journal of Hand Therapy* 14.2, pp. 128–142.

Bellamy, Nicholas, W Watson Buchanan, Charles H Goldsmith, Jane Campbell, and Larry W Stitt (1988). "Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee." In: *The Journal of rheumatology* 15.12, pp. 1833–1840.

DASH (2006). *The Disabilities of the Arm, Shoulder and Hand Outcome Measure*. Accessed: 2025-11-06. URL: http://www.dash.iwh.on.ca.

Dawson, Jill, Ray Fitzpatrick, David Murray, and Andrew Carr (1998). "Questionnaire on the perceptions of patients about total knee replacement". In: *The Journal of Bone & Joint Surgery British Volume* 80.1, pp. 63–69.

OKS (2016). *The Oxford Knee Score*. Accessed: 2025-11-06. URL: https://innovation.ox.ac.uk/outcome-measures/oxford-knee-score-oks.

21. Extended Reality for Safety and Social Interaction at Work

Leesa Joyce[1], *Pablo Rocamora*[2], *Blanca Guerrero*[3], *Jordi Almendros*[3], *Federica Bertolucci*[4], *Theodora Pistola*[5], *and Fabio Carrara*[6]

[1] HOLO-Industrie 4.0 Software GmbH, Germany,
[2] FACTOR, Spain,
[3] Research Centre on Production Management and Engineering (CIGIP), Universitat Politècnica de València (UPV), Spain
[4] Azienda USL Toscana Nord-Ovest (ASL-NO), Italy
[5] Centre for Research and Technology Hellas, Information Technologies Institute (CERTH), Greece
[6] Institute of Information Science and Technologies, National Research Council (CNR-ISTI), Italy

Abstract. SUN partners have implemented a pilot project at FACTOR, a metal parts manufacturing facility, to explore how XR technologies can improve safety, social interaction, and ease task management in industrial settings. The pilot, titled "Extended Reality for Safety and Social Interaction at Work", encompasses two complementary use cases. The first focuses on safety training and practice through immersive XR experiences, while the second targets real-time task prioritization and workflow optimization on the shop floor. In use case 1, workers use the SUN-SET application through the AR headset (HoloLens2) to engage in interactive training on Personal Protective Equipment (PPE) guided by an avatar, followed by live observation tasks in real workplace environments supported by object recognition and real-time visual feedback. Use case 2 introduces the PRIORI-XR algorithm, which dynamically assigns and updates work priorities based on sensor and camera data, displaying tasks directly through the XR interface to support efficient decision-making and reduce cognitive load. The system architecture integrates client–server modules, AI-based image recognition, hand gesture interaction, and cybersecurity monitoring, enabling robust and adaptive XR functionality across both cases. Validation of the technology was conducted in two phases—initially with SUN project staff and subsequently with factory operators in real environments. User feed-

back highlighted high acceptance, comfort, and perceived usefulness, particularly in training engagement and workflow organisation. Challenges related to gesture precision and detection stability were progressively resolved through iterative improvements. Overall, the pilot project demonstrates the feasibility and value of XR as a human–machine interface in industrial contexts, enhancing both safety culture and operational efficiency.

21.1 Introduction

The integration of Extended Reality (XR) technologies into industrial environments is transforming the way manufacturing industries function. The current research focused on investigating how XR can enhance both workplace safety and productivity at manufacturing facilities where safety and productivity play crucial roles. The pilot "Extended Reality for safety and social interaction at work" was divided into two distinct yet complementary scenarios: use case 1 – Safety Training and Practice through XR, and use case 2 – XR-based Task Prioritisation in Industrial Settings. Both cases leverage the HoloLens 2 headset, the SUN-SET application, and a range of supporting systems, including object recognition, gesture-based interaction, and decision-making algorithms. The overarching goal of the pilot is to examine how immersive technologies can contribute to safer, smarter, and more human-centered workplaces by combining intelligent data processing with natural and intuitive user interfaces. The use cases were validated at FACTOR, a manufacturing industry that is a leading producer of complex metal parts.

Use case 1 focuses on enhancing PPE-related training and compliance through interactive and data-driven XR modules. Using the HoloLens 2 integrated with the SUN-SET system, users engage in guided training where an avatar delivers multimodal instructional content—text, images, and videos—followed by comprehension assessments. The system automatically generates a digital protocol documenting user performance, which is securely transmitted for traceability and verification of safety compliance. In the extended real-world subcase, participants move through the workplace evaluating PPE compliance of employees across different roles. Using a speech command, they capture images that are instantly analyzed by both the user and an object recognition tool, with results cross-verified to generate feedback reports. This allows real-time learning and correction in practical operational contexts.

Use case 2 introduces PRIORI-XR, an adaptive XR-based task management framework for dynamic prioritization in industrial environments. Through continuous data acquisition from distributed sensors and cameras, the system evaluates real-time operational states—such as machine faults or container levels—to compute and display ranked task lists directly in the HoloLens 2 interface. The worker can interact with this list to accept, defer, or reassign tasks through hand gestures while the algorithm autonomously updates priorities in response to changing conditions. The system also incorporates multi-user coordination logic to resolve task allocation conflicts, ensuring synchronized and efficient workflow execution.

21.2 Methods and Results

The implementation of the pilot at FACTOR has been structured around the two defined use cases, each of which relies on a modular architecture that integrates components serving different functionalities to enhance the AR application SUN-SET. The SUN-SET application has 2 components: the client side and the server side. The entire application is remotely rendered on a server (enhancing the computational capacity) and is streamed to the client HoloLens 2. The user input and the SLAM (Simultaneous Localization and Mapping) data are streamed from the client to the server. The SUN-SET application integrates multiple components relevant to specific use cases.

At the core of the server bundle, the SUN-SET Server hosts the training and practice applications for use case 1 and manages the integration of several modules, including the Avatar and the object recognition tool (see Figure 21.1). The Avatar is responsible for improving the visualization of training content and guiding the user with directives or prompts throughout the session. The object recognition tool (see Chapter 5) processes frames from the HoloLens cameras to detect PPE on external actors during practice sessions. The frames are transmitted from the SUN-SET Server to the object recognition tool, while the detection results are sent back to the server. This process allows the headset to provide users with real-time feedback in the form of visual overlays highlighting whether safety equipment is being correctly worn. Additionally, the Tokenized Platform (see Chapter 17) supports the entire architecture by storing system data such as training protocols, results, and multimedia content, while also enabling asset ownership and access control.

Use case 2 focuses on task prioritisation, and integrates both physical and digital components to deliver real-time decision support to operators on the shop floor (Figure 21.2 reports the use case 2 architecture). For this implementation, two outdoor cameras were installed at different heights within the FACTOR shopfloor to monitor the state

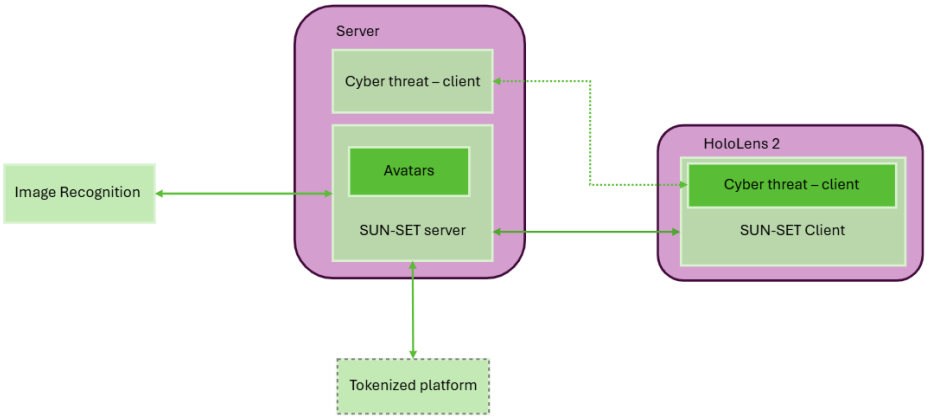


Figure 21.1: Use case 1 - Architecture.

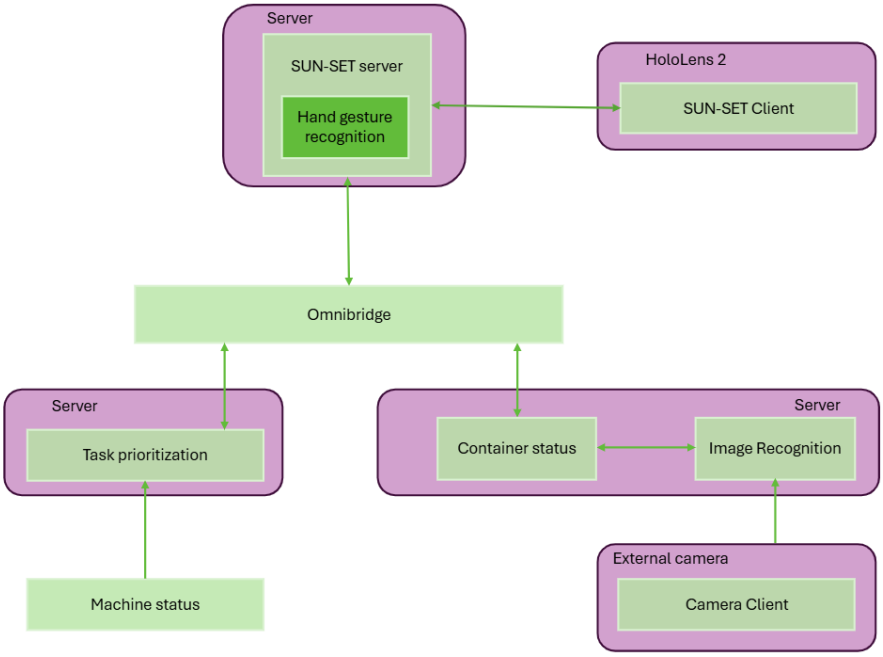


Figure 21.2: Use case 2 - Architecture.

of raw material and waste containers attached to the machines. The camera feeds are analyzed by the object recognition tool mentioned above, which perceives the fullness status of waste and raw material containers. This provides input for the PRIORI-XR algorithm, a rule-based logic system designed to automate and optimise task management in industrial environments (see Chapter 10). The algorithm begins by loading data on machines, containers, production status, and pending tasks. It classifies the current shift, filters completed tasks, and analyses operational needs such as empty material containers or full waste containers, generating replenishment and emptying tasks accordingly. It also assigns cutting tasks to sawing machines when capacity allows, and during shift changes, it automatically creates tasks for cleaning and maintenance. Each task is assigned an identification, description, priority level, and a status that begins as "Pending" and can subsequently be changed to "Accepted", "Rejected", or "Completed" by the operator through hand gestures. The hand gesture recognizer adds interactivity to the training process by interpreting gestures such as thumbs-up, thumbs-down, and two-finger swipe, which the user can employ to respond to prompts without the need for traditional controllers (see Chapter 14. Communication with external modules is handled by the OmniBridge component (see Chapter 16), which serves as a unified messaging broker, ensuring smooth and reliable data exchange between the server and external services. If a task is rejected, the system automatically reactivates it as pending, ensuring that no critical operation is left unresolved. Two key components integrated in the task prioritization use case are the container flow module and the factory data integration. The container flow module enables real-time tracking of raw material usage and waste generation at the machine level. It integrates with machine status data to monitor production progress. When specific thresholds are reached, the module automatically triggers corresponding tasks. Its event-driven logic ensures prompt task generation, even in zones lacking camera coverage, by leveraging the factory's internal data infrastructure. This approach improves responsiveness and ensures continuous operations where visual monitoring is not feasible. The operator experiences this workflow and the pending tasks directly through the HoloLens 2, where the task list is displayed in Spanish with clear priority labels.

21.3 Validation

The validation of Pilot 2 was conducted in two main phases: an initial experimental phase involving volunteers from the SUN staff, and a second phase involving operators from FACTOR in Valencia, who participated as end users during a regular working day.

Phase 1: Experimental Validation

In the first validation phase, both use case 1 and use case 2 were tested in a dedicated room at FACTOR.

For use case 1, PPE detection, seven participants were equipped with the HoloLens 2 AR headset and engaged in an interactive training session designed to train and identify the appropriate PPE—specifically, glasses, gloves, and helmets—required in specific areas of the shopfloor. The training involved reading instructions and viewing accompanying images illustrating PPE compliance. At the end of the session, users answered a series of questions assessing their understanding. Interaction with the system was achieved through the selection of virtual buttons appearing in front of the user. Throughout the process, an avatar of a worker guided the user, providing positive or negative feedback depending on the correctness of their responses. Following the training, participants used the XR glasses in pairs (one with and the other without PPE) to verify correct or incorrect PPE usage. The system provided immediate visual feedback in the form of green or red indicators (bounding boxes) corresponding to the type of equipment correctly or incorrectly worn.

For use case 2, shopfloor safety and object tracking, users wearing the XR glasses were presented with a predefined list of prioritized and assigned tasks intended to support efficient task distribution per shift. Users indicated to the XR system whether each task had been completed. Task prioritization was done automatically using real-world data collected from cameras on the shopfloor, which detected the levels of raw materials and waste containers.

At the end of the session, all participants completed the user questionnaire, a five-point Likert scale survey (21 multiple-choice questions ranging from "Strongly Agree" to "Strongly Disagree"). The questionnaire assessed system usability, device wearability, system reliability and performance, learning and skill development, and overall satisfaction. The overall user feedback was positive. Participants found the device comfortable to wear and considered the training proposed in use case 1 more engaging than standard training. Regarding use case 2, users reported that the system could be highly useful for task management. The user feedback included suggestions to improve the system with easier UI, and there were requests to include auditory feedback to improve recognition of correct or incorrect PPE use. Some participants initially experienced difficulty performing the "pinching" gesture used to select tasks; however, they reported that interaction became intuitive after brief practice.

Phase 2: Real-Environment Validation

In the second validation phase, user feedback from the initial testing was incorporated to improve the platform's usability and facilitate smoother interaction in real-world conditions. Eleven workers from FACTOR participated in this phase.

Figure 21.3: Use case 1 validation sample for PPE detection.

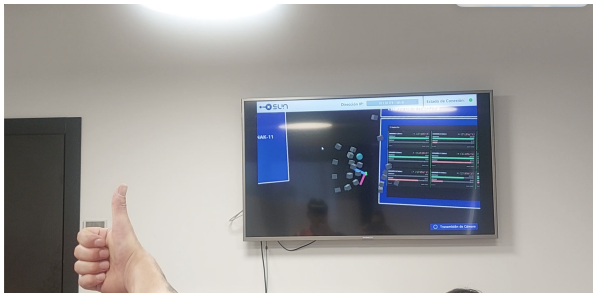


Figure 21.4: Use case 2 validation sample for shopfloor safety and object tracking.

For use case 1, the users were initially trained through virtual training similar to what has been done in phase 1 regarding the appropriate PPE for specific areas in the shopfloor. Later, participants moved with the AR headsets to the real environment where employees naturally worked in different areas. Participants were asked to verify the correctness of PPE usage by other factory workers across three operational contexts: logistics personnel, machine operators, and crane operators. Participants could capture photographs of the worker in front of them through their headsets by giving a speech command ("Photo"). The picture was shown to the user immediately, with the option to report whether the worker used the right PPE for the specific area. At the same time, the system automatically evaluated the photograph for the correctness of PPE usage through the object recognition tool (as shown in Figure 21.3. The system then verified the user's answers with its detection to generate a report that helps the user recognize faulty evaluations. The object recognition tool functioned like a personalized trainer.

In use case 2, participants received prioritized tasks. The users could accept, reject, or report completion through hand gestures (thumbs up, thumbs down, and two-finger swipe, respectively, see Figure 21.4 for an example). After completion, users could

report the number of raw materials filled on the dashboard in the virtual application. The dashboard also allowed users to visualise detailed machine status.

All participants completed the user questionnaire, and the analysis of responses indicated high levels of system acceptability and user satisfaction. Participants reported that the device was comfortable to wear and suggested that publishing the task prioritization panel in the support area would help organize workflows more effectively. While some participants found the "pinching" gesture initially difficult, they confirmed that interaction became straightforward after practice. They also noted that the hand gesture was easier and more efficient for interacting with the interface. Although still at an early stage of development, the system was well received by users and shows strong potential for application in real industrial environments, provided that its operation is made more agile and seamless.

21.4 Conclusions

This pilot project at FACTOR demonstrated how XR can be effectively applied to enhance both safety and productivity in industrial environments through immersive interaction and intelligent automation. Use case 1 showcased how XR can transform traditional safety training into an engaging, traceable, uniform, and measurable process that seamlessly transitions from virtual instruction to real-world application. The combination of avatars, hand gestures, and real-time PPE detection promotes active learning and reinforces safety culture. Use case 2 extended the application of XR to operational decision-making, enabling real-time task prioritisation based on contextual data from visual and sensor inputs. The PRIORI-XR algorithm and the integration with data infrastructures highlight the feasibility of combining XR with industrial IoT and AI for smart manufacturing.

Together, these cases underline the potential of XR not only as a visualization tool but as a core component of human–machine collaboration. By reducing cognitive load, improving situational awareness, and fostering engagement, XR technologies can bridge the gap between digital intelligence and human intuition, paving the way for safer, more adaptive, and more efficient industrial environments.

22. Extended Reality for People with Serious Mobility and Verbal Communication Diseases

Vincent Mendez[1,2], *Aiden Xu*[1], *Simona Losacco*[1], *Elena Ferrazzano*[1], *Federica Bertolucci*[3], *Federico Posteraro*[3], *Leesa Joyce*[4], *Ioannis Paraskevopoulos*[5], *Ferdinando Bosco*[6], *Leonardo Corsano*[6], *Bertrand Léger*[7], *Daniele Leonardis*[8], *Spyridon Symeonidis*[9], *Sotiris Diplaris*[9], *George Loukas*[10], *Riccardo Bovo*[10], *Friedhelm Hummel*[1], *and Silvestro Micera*[1]

[1] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[2] Centre Hospitalier Universitaire Vaudois, Switzerland
[3] Azienda USL Toscana Nord-Ovest (ASL-NO), Italy
[4] HOLO-Industrie 4.0 Software GmbH (HOLO), Germany
[5] ThinGenious PC (THING), Greece
[6] Engineering Ingegneria Informatica S.p.a. (ENG), Italy
[7] Department of Medical Research Suva Clinics, Clinique Romande de Réadaptation (CRR-SUVA), Switzerland
[8] Scuola Superiore Sant'Anna (SSSA), Italy
[9] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece
[10] University of Greenwich (UoG), UK

Abstract. Individuals with severe motor impairments and clinical apathy face significant barriers to communication and social interaction. This chapter presents an Extended Reality (XR) platform designed to address these challenges by enabling users to interact within a virtual environment using non-invasive body-machine interfaces. The system translates residual forearm muscle activity, captured via electromyography (EMG), into avatar control, while integrated haptic and thermal feedback enhances immersion and interaction fidelity.

An initial technical validation with healthy volunteers confirmed the platform's stability and high user acceptance. The multisensory feedback was shown to be

particularly effective in increasing the sense of presence. Furthermore, an integrated cybersecurity module was successfully validated.

A subsequent clinical feasibility study with patients affected by neurological disorders such as stroke, spinal cord injury, and mild cognitive impairment demonstrated that the platform is usable by the target population and supports both immersive interaction and the development of an innovative intervention for apathy. This intervention combines a VR-based effort decision-making task with non-invasive transcranial temporal interference (TIS) stimulation to target the reward system. Overall, this work establishes a robust framework for multisensory XR systems, offering a promising modality for interaction, rehabilitation, and motivation enhancement in individuals with severe disabilities.

22.1 Introduction

The focus of Pilot 3 was to apply the integrated SUN platform to address the complex needs of individuals with severe neurological disorders. The validation was structured around two distinct clinical scenarios:

- *Case 1: Immersive Interaction for Severe Motor Impairments.* This scenario was designed for individuals with conditions such as incomplete tetraplegia (Spinal Cord Injury) or stroke. The goal was to deploy a system enabling them to interact within a personalized virtual environment using non-invasive body-machine interfaces (EMG) and multisensory feedback (haptics, thermal);
- *Case 2: Addressing Clinical Apathy.* This scenario tested a virtual reality adaptation of a well-established effort-based decision-making task to objectively assess patients' motivation. The aim was to validate the XR platform's implementation, showing increased engagement and immersion compared to the lab-based version of the task, and to pave the way for the development of a VR-based intervention for apathy based on non-invasive brain stimulation of the reward system during task performance.

This chapter details the final integrated architecture, experimental methodology, and validation results from the clinical feasibility studies conducted for both of these use cases.

22.1.1 The Challenge: Overcoming Physical and Motivational Barriers

Severe motor and communication impairments, resulting from conditions such as incomplete tetraplegia, stroke, or cerebral palsy, drastically limit an individual's ability to interact with their environment and maintain social connections, often leading to significant psychological strain. For instance, a person with tetraplegia may be confined to a clinical setting, longing for the freedom to interact with loved ones in familiar environments beyond the clinic walls.

A distinct but equally debilitating challenge is clinical apathy, a disorder of motivation characterized by diminished goal-directed behaviour, which is highly prevalent in various neurological and psychiatric disorders, including traumatic brain injury, Parkinson's and Alzheimer's disease, and stroke. Apathy represents a major barrier to rehabilitation, as it diminishes patients' willingness to engage in therapeutic activities, including the adoption of new assistive technologies. Therefore, developing targeted interventions to alleviate apathy is essential to enhance motivation and promote engagement with rehabilitation.

22.1.2 The Vision: An XR Platform for Immersive and Meaningful Interaction

To address these challenges, we have developed an XR platform that applies a new generation of non-invasive bidirectional body-machine interfaces (BBMIs). These interfaces are designed to allow a smooth and very effective interaction for people with different types of sensory-motor disabilities within a VR environment. The goal is to extend the traditional boundaries of interaction by mapping residual muscular activities to a spectrum of actions within a familiar virtual setting.

This system is designed to improve user function and immersion by integrating multisensory feedback channels. Haptic and thermal stimuli are systematically employed to create a high-fidelity virtual environment that more accurately simulates physical interaction. The objective is to provide users with an effective alternative modality for communication and environmental control, thereby mitigating some of the functional deficits imposed by their physical condition.

22.1.3 Key Use Cases

The platform's capabilities were channelled into two primary use cases, each targeting a specific aspect of the overall challenge.

Use Case 1: Immersive Interaction for Individuals with Limited Mobility

This scenario is designed for individuals with severe upper limb disabilities. It provides the ability to interact with loved ones and objects in a personalized virtual space by utilizing electromyography (EMG) signals to capture residual muscular activity from the forearm. These signals are decoded to control an avatar's navigation and hand gestures, enabling intuitive interaction without conventional controllers.

To deepen the sense of presence, the visual feedback within the VR environment is augmented with multisensory feedback. A wearable haptic device provides modulated clenching and vibrotile cues to inform the user which action has been decoded from the EMG system. Furthermore, thermal feedback is integrated to enhance the realism of interactions, allowing users to experience temperature sensations, such as touching a cold glass or feeling the warmth of a simulated human touch. For patients unable to speak, the VR application also includes a communication interface that allows them to construct sentences by selecting predefined syntactic elements. This integrated approach offers a novel and rich communication pathway for those with severe motor and verbal disabilities.

Use Case 2: Addressing Clinical Apathy through XR

This scenario focuses on testing a virtual reality (VR) implementation of a well-established effort-based decision-making task, in which patients decide whether to exert physical effort in exchange for rewards. This paradigm has been widely used to study apathy, defined as a reduction in goal-oriented behavior, which manifests as reduced motivation. The VR adaptation expands the range of rewards beyond classic monetary incentives to include interaction with a relative's avatar or the opportunity to watch an enjoyable video. By analyzing participants' choices and decision times, it is possible to objectively quantify motivational state and assess the subjective value assigned to specific offers. Compared with the traditional lab-based version, the VR implementation enables assessment of motivation in a more realistic yet clinically compatible environment. The ultimate goal is to develop an innovative therapeutic intervention based on transcranial temporal interference stimulation, which allows targeted, non-invasive modulation of deep brain regions. Participants receive this stimulation concurrently while performing the task, aiming to enhance motivation by modulating neuronal plasticity in reward-related circuits.

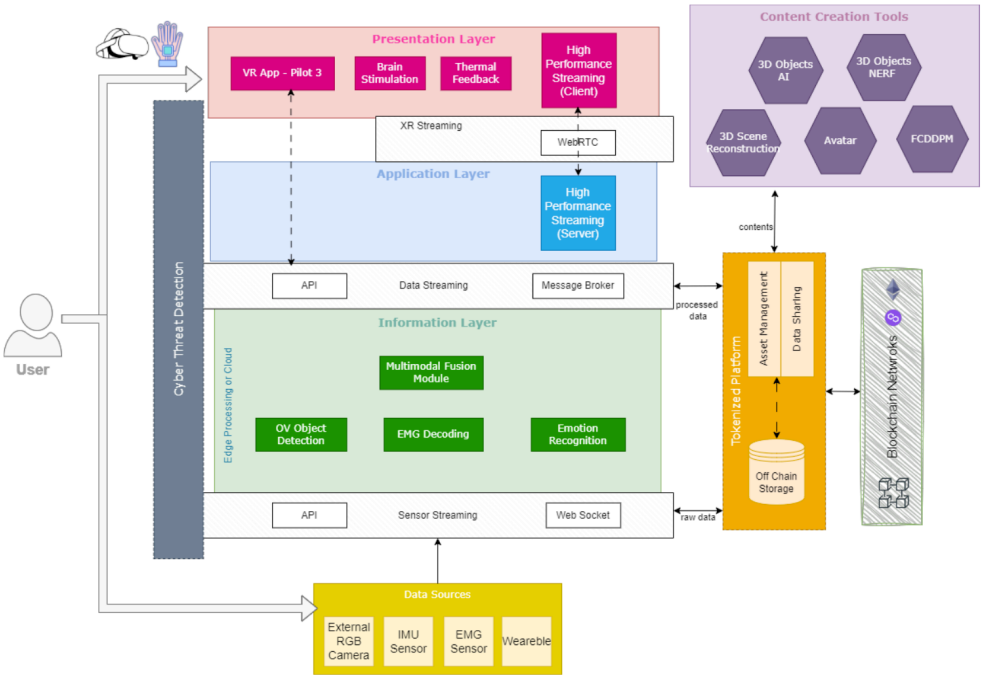


Figure 22.1: High-level architecture for the pilot, illustrating the flow of data from sensor sources through the information and application layers to the user presentation layer. Figure adapted from SUN Deliverable D6.2.

22.2 Methodology and Results

The development and evaluation of the SUN platform followed a structured approach, encompassing system design, component integration, and a formal two-stage validation process.

22.2.1 System Architecture and Implementation

The platform is built on a modular architecture designed for real-time, multisensory XR experiences (see Figure 22.1). The core of the system is a VR application developed in the Unity game engine. The architecture was tailored to the specific needs of the two use cases.

Components for Use Case 1 (Immersive Interaction)

This architecture required the integration of multiple custom hardware and software modules to enable non-invasive control and feedback:

- *VR Application:* Developed using Unity, the application manages the virtual environment, multi-user sessions (via Photon), and real-time voice communication (via Vivox). It supports multiple control modalities, including hand-tracking, gaze-based selection, and EMG control;
- *EMG Decoding:* User intent is captured via an EMG system that records electrical activity from the forearm muscles. A classifier-based model was trained to decode gestures (e.g., wrist flexion/extension), which were mapped to navigation and interaction commands (see also Chapter 11);
- *Multisensory Feedback Systems:* To enhance immersion, two distinct systems were integrated: (1) a **haptic armband** providing vibrotactile feedback to confirm decoded EMG actions, and (2) a **thermal feedback** device (refined to be wireless in the clinical validation) to simulate temperature when touching virtual objects (see also Chapter 6);
- *Emotional state tracking:* Physiological signals, including photoplethysmography, heart rate, electrodermal activity, and temperature, were measured using the EmotiBit wearable device. The recorded physiological signals were utilized to infer the emotional state of the participant during the experiment in real-time (see also Chapter 15);
- *System Integration:* The VR app, EMG system, and thermal feedback device communicated via WebSocket connections, while the haptic system used a TCP/IP connection.

Components for Use Case 2 (Apathy Protocol)

This architecture was designed to implement a robust effort-based decision-making task, comparing a traditional setup with an immersive VR one.

- *Effort-based decision-making Task Software:* This task, based on established protocols to characterize goal-oriented behavior, was developed in two forms: a standard 2D application and an immersive VR application in Unity;
- *VR Environment:* In the VR version, the virtual scene consisted of two interconnected rooms separated by a sliding door. Within this environment, offers were

presented to participants, with the effort component represented by the degree to which the door must be opened, and the monetary reward component represented both numerically and visually—displayed on a virtual tablet next to the door and reflected in the number of coins visible inside a cabinet. This design replaced the previous gauge-based representation of effort (bar height) with a more natural, embodied interaction that connects physical action and visual feedback in an intuitive way;

- *Hardware Integration:* The system used a standard keyboard as the input device. Participants exerted physical effort by repeatedly pressing the 'Ctrl' key. Different effort levels were defined as percentages of each patient's maximum tapping frequency, measured during a prior calibration phase. Participants wore a VR headset to experience the task in an immersive virtual environment. For the clinical validation, non-invasive transcranial Temporal Interference Stimulation (tTIS) electrodes were positioned on the patient's scalp to verify the feasibility of simultaneous electrode placement and headset use; however, stimulation was not applied during this phase.

22.2.2 Validation Experiments

The platform's evaluation was conducted in two stages: an initial technical validation with healthy volunteers, followed by a clinical feasibility validation with the target patient population.

Technical Validation with Healthy Volunteers

The first validation was conducted at the EPFL Campus Biotech in Geneva. The primary objectives were to achieve full integration of all system components for Use Case 1 and to test its usability with healthy subjects.

Six healthy volunteers from the SUN consortium participated. Participants wore a Meta Quest VR headset, an EMG sleeve, and the haptic and thermal feedback devices (see Figure 22.3). Participants were asked to complete a sequence of tasks within the virtual room (see Figure 22.2) using the EMG-based control system, including navigation and interaction with thermal objects. Each successfully decoded action was confirmed via haptic feedback.

Clinical Feasibility Validation with Patients

The final validation was conducted at the Clinique Romande de Réadaptation (CRR) in Sion, Switzerland, with patients for both use cases.

(a) (b)

Figure 22.2: Side-by-side views of the VR application: (a) Interactive scene: the objects are placed on a desk and consist of a pumpkin and a mug that can be interacted with. (b) Screenshot of the menu when approaching the desk, showcasing the four different actions with the correct hand gesture.

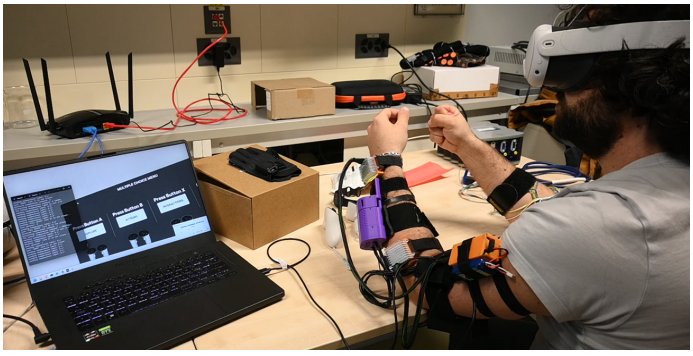


Figure 22.3: A participant during the technical validation session wearing the full set of components for Use Case 1. Figure adapted from SUN Deliverable D6.2.

- *Use Case 1 (Immersive Interaction):* Two patients with upper limb impairments (one post-stroke, one with incomplete spinal cord injury) participated. After calibration of the EMG system, each patient performed a guided session to navigate the virtual room and interact with objects using EMG control. This was followed by a comparative session where the same tasks were performed using eye-tracking for control;
- *Use Case 2 (Apathy Protocol):* One patient with mild cognitive impairment participated. The protocol began with a calibration phase to measure participants' maximum tapping frequency, which was then used to set the effort levels as percentages of this maximum. The patient subsequently completed two blocks of the classic lab-based effort-based decision-making task, followed by two blocks in the VR environment, enabling a direct comparison between the two conditions.

22.2.3 Experimental Findings

The validation process provided key insights into the system's performance, usability, and clinical feasibility.

System Usability and User Experience

The initial technical validation with healthy volunteers demonstrated stable integration of all core components. Participants reported a high degree of immersion and satisfaction, with particularly positive feedback on the multisensory feedback. The thermal and haptic cues were found to significantly enhance the sense of presence and interaction realism.

While the EMG control was functional, some users found it less intuitive than a gaze-based alternative for menu selection, though it was preferred for environmental navigation. Several participants noted that the physical setup (particularly the wired thermal device) could be cumbersome, highlighting the need for a more streamlined, wireless hardware configuration that was successfully developed and deployed in the subsequent clinical validation.

Therapeutic Value and Patients' Engagement

Feedback from patients was collected through specific questionnaires concerning the usability of the system, the effectiveness of the various haptic feedbacks, the effectiveness of the virtual interface, emotional engagement, overall experience, and also some technical aspects of the sessions. In use case 1 (Immersive Interaction), both patients reported that the setup was comfortable, that they did not feel fatigued after

the session, and that they felt comfortable with both tracking systems (eye tracking and EMG tracking), but the eye tracking system seemed more efficient. The temperature feedback was reported as realistic, but not the haptic one. The interaction with the avatar was not found to be realistic. Nevertheless, the overall experience was positive, and both patients were satisfied after the session. Feedback was also collected from clinical staff, in particular from two physicians specialized in rehabilitation who attended the sessions. From their side, evaluations were positive, especially in view of future developments aimed at improving avatar customization and the ease of use of the virtual interface. In Use Case 2 (Apathy Protocol), the patient completed a specific questionnaire comparing the VR-based session with the traditional lab-based session. The analysis of the responses indicated that the participant found the VR environment realistic and did not experience nausea or post-session fatigue. He also reported that the headset was comfortable and did not interfere with task performance. Overall, the patient expressed satisfaction with the VR session. Although the lab-based version was perceived as more intuitive, the participant found it easier to maintain attention during the VR session and felt capable of completing a greater number of task blocks in a single day using the VR setup. Moreover, the patient reported that the VR version would increase his willingness to attend sessions on multiple days. This was considered advantageous, as an intervention for apathy based on transcranial temporal interference stimulation would require participants to complete multiple task blocks within a single session, with sessions repeated over several days. Overall, the participant regarded the VR-based protocol as a useful and engaging alternative to the conventional lab-based task.

Cybersecurity Validation in an XR Environment

A separate experiment with 15 participants was conducted to validate the platform's integrated Cyber Threat Detection module. In this study, a GPU-based malware attack was simulated during the XR session, causing performance degradation (e.g., frame rate drops). An Intrusion Detection System (IDS) was designed to detect this attack and present a multimodal (audio-visual) warning to the user.

- *User Response:* The security warning proved highly effective. All participants either acted on the warning immediately or sought confirmation from the experimenter, with none ignoring it. Questionnaire results confirmed that the warning was perceived as attention-grabbing, clear, and actionable;
- *Emotional Response:* Using the Self-Assessment Manikin (SAM), results showed that the warning induced significantly higher arousal, indicating it was highly effective at capturing attention, but did not trigger a strong negative or positive

emotional valence. This suggests the alert was perceived as an informative, actionable prompt rather than a distressing one;

- *IDS Performance:* The IDS demonstrated high accuracy, with a 0% false positive rate across all sessions. The average detection latency for the GPU-based attack was 1.25 seconds, indicating a prompt and reliable response.

22.3 Conclusions

The work presented in this chapter demonstrates the successful implementation and full validation of a novel XR platform designed to enhance interaction for individuals with severe motor and communication impairments. The validation process, conducted first with healthy volunteers and culminating in a clinical feasibility study with patients, confirmed that the platform is robust, usable, and well-received.

The initial technical validation was crucial for confirming the stability of the integrated system and affirming that multisensory (haptic and thermal) feedback significantly contributes to a realistic sense of presence. This phase also identified key hardware limitations, such as the cumbersome wired setup, which were successfully addressed with the development of a wireless thermal device.

The subsequent clinical validation at the CRR in Sion successfully translated the platform from the lab to the target patient population. This study confirmed the clinical feasibility of deploying the full system for patients with stroke and spinal cord injuries. It also provided a direct comparison between EMG and eye-tracking control modalities, yielding essential insights for designing future adaptive and user-centric interfaces. For the apathy use case, the validation demonstrated that the XR platform provides a viable and engaging environment for implementing therapeutic interventions combining VR with non-invasive brain stimulation, during concurrent performance of an effort-based decision-making task.

In summary, this pilot study established a solid foundation and a clinically validated framework for employing multisensory XR as a powerful tool to support individuals with severe disabilities. Future work will focus on longitudinal studies to evaluate the platform's long-term impact on quality of life, social interaction, and—specifically for Use Case 2—the alleviation of apathetic symptoms when combined with an active tTIS protocol. Further refinement of hybrid EMG–gaze control systems will also be prioritized based on insights gained from these studies.

Sustainability, Ethics, and Impact

The concluding Chapters of the SUN Book present a comprehensive perspective on the project's evolution from innovative research to real-world impact. The results achieved through clinical validation, advanced technological development, and ethical deliberation merge into a unified strategy that ensures long-term sustainability. The SUN ecosystem stands as a model for the effective translation of human-centred AI and XR technologies into practical healthcare and well-being solutions. Its exploitation strategy is designed to generate value for patients, clinicians, and industry stakeholders alike, facilitated by shared platforms, a commitment to open innovation, and the creation of new business opportunities that reach beyond the scope of the initial pilots. At the same time, the project evidences a strong commitment to social and ethical responsibility. It meets transparent data governance, consistent compliance with European legal frameworks, and respect for privacy and human dignity. By integrating technical excellence with ethical integrity and strategic market insight, SUN establishes a robust foundation for a new era of responsible, sustainable, and impactful digital health innovations.

23. Exploitation and Business Model

Alexandru Stan[1] *and George Ioannidis*[1]

[1] IN2 Digital Innovations GmbH, Germany

Abstract. This chapter presents the exploitation strategy developed within the SUN project, which aims to bring human-centered Extended Reality (XR) technologies from research prototypes to sustainable real-world applications. It outlines the methodologies used to identify and mature project results into Key Exploitable Results (KERs). The KERs underwent targeted business development, for some with the help of the Horizon Results Booster programme, using tools such as the Value Proposition Canvas and the Business Model Canvas to define clear value propositions and exploitation pathways. The chapter situates these efforts within a rapidly expanding XR market, highlighting Europe's competitive advantages in industrial and healthcare applications. It details several exemplary KERs, including the SUN Integrated Platform and novel XR interfaces. Overall, the SUN exploitation methodology demonstrates how structured innovation management and open collaboration can bridge the gap between research and market impact in next-generation XR ecosystems.

23.1 Introduction

Research and innovation projects like SUN are complex undertaking where technologies and methodologies from a certain field are brought through targeted research activities beyond the state of the art, are combined with other beyond the state of the art technologies, and integrated into larger systems that are then validated for the first time with users. Maturity levels of technologies involved are often mapped to a Technology Readiness Level (TRL) [APRE and CDTI 2022]. In this way, the technologies and methodologies developed are brought from the level of a proof of concept

(TRL 3) or components validated previously in laboratory environments (TRL 4) to a more technologically mature level such as TRL 5, if the technological components are integrated for testing in a simulated environment, or TRL 6, if the technological components are integrated in a prototype that is close to the desired configuration and is tested in a simulated operational environment. Indeed, SUN has brought many different technologies to a demonstration level maturity (TRL 5/6).

Many European-funded research projects successfully demonstrate advanced technological prototypes, reaching TRL 5 or 6. However, a significant challenge remains to bridge the gap — often called the "valley of death" — to bring these innovations to market as viable products or services (TRL 8/9) [Hirzel et al. 2018]. This challenge is particularly acute for complex systems like those developed in SUN, which combine novel hardware, artificial intelligence, and different software ecosystems. Without a clear and early-stage strategic plan, valuable research outcomes risk remaining as successful demonstrations with limited real-world socio-economic impact.

The SUN project was conceived to address this challenge directly. Our core mission is not only to investigate and develop next-generation XR solutions but also to ensure that they are fundamentally plausible, interactive, and centered on human social needs. This human-centric focus, applied to domains like physical rehabilitation, industrial safety, and assistive communication, imbues the project with a clear societal mandate. Consequently, a robust exploitation strategy is not just an afterthought but a foundational pillar of the project, essential to translate our research achievements into tangible benefits for researchers, citizens, clinicians, and workers.

This chapter details the multi-faceted exploitation strategy designed and executed within the SUN project. It serves as a blueprint for transforming our research outcomes from demonstration to deployment. We begin by defining the methodology planned for exploitation planning and by having an overview of the market and business landscape. Following this, we identify and categorize the project's Key Exploitable Results (KERs) and present some insights into the planned exploitation roadmap for them. Finally, we conclude by outlining the hybrid business and exploitation models chosen to maximize impact. This structure provides a comprehensive roadmap for post-project sustainability and adoption of the SUN results.

23.2 Methodology for Exploitation Planning

23.2.1 What is Exploitation in the context of EC-funded projects

A central objective for any Horizon Europe project is to make sure that the research and demonstration carried out in the project can lead to towards tangible, long-term impact. All projects must strive to transform project results into concrete benefits for society, maximising the scientific, social, economic, technological, and policy value of the framework programme under which they have been (co-)financed. This process is formally defined by the European Commission as "exploitation", which constitutes the use of project results in one of four primary, non-exclusive pathways:

- *Further Research*: Using results in subsequent research or innovation activities;
- *Commercialisation*: Using results in developing, creating, or marketing a product, process, or service;
- *Standardisation*: Using results to inform or create new technical standards;
- *Policy Making*: Using results to provide evidence and inform public policy.

Thus, exploitation should be understood as how to mobilise outcomes and achieve impact thanks to long-term sustainability planning. The key aspect when discussing *exploitation* is the **use of results**[1], "translating research concepts into concrete solutions that have a positive impact on the public's quality of life". This is why exploitation can be understood as a *value-driven process*, where *value* can be understood in several ways:

- Generate revenues (e.g., through commercial use with paying customers);
- Fulfilling an existing market or societal gap (e.g., not for profit, providing better services, improved delivery processes, support for policies);
- Increase the intangible assets in the organisation or in the community (e.g., distinctive skill set, standards, etc.).

It is important to note from the above that when speaking about exploitation, this is about making concrete use of the achieved research results without restricting the activities to commercial use.

[1] https://research-and-innovation.ec.europa.eu/strategy/dissemination-and-exploitation-research-results_en

The SUN project's exploitation strategy was designed as a systematic process to identify, mature, and prepare project results for all potential exploitation pathways. This required a clear methodology for first identifying *what* to exploit (i.e., key exploitable results) and then *how* to plan for its uptake (e.g., further R&D, PhD thesis or postdocs, patent, licensing, establishment of a spin-off or start-up, manufacturing, direct sales, service provisioning, training, educational activities, policy making).

23.2.2 Identification of Key Exploitable Results (KERs)

To structure this process, the first step is to distinguish general "results" (any tangible or intangible output) from **Key Exploitable Results (KERs)**. A KER is a result formally assessed by its owners and the consortium as having **significant potential for exploitation**, be it scientific, commercial, or societal. It is the primary unit around which a concrete exploitation plan is built.

The BOOSTER programme [Commission 2020] encourages to thinking of KERs as *special results* that respond to the specific needs of a well-defined group, the *adopters*, and solve their needs much better than "state of the art". The KER could be a product or a process, a format for a new service, a new standard, a new training course, or even input for a new project. It is up to the partners to decide which of the project results they choose for **use** and/or **market introduction**. The *use* is understood as "the action made by adopters to make the KER available to others". It is important to note at this point that the *use* can be both **commercial** and **not for profit** (such as improving a policy action or public knowledge).

Moreover, it is important to note that partners can choose to exploit the results in two different ways:

- *Directly:* partners exploiting the results themselves. For low-TRL results, this can materialise through the results becoming background knowledge in future research activities, while for high-TRL activities, partners often choose to develop and commercialise a product or process. Moreover, another type of direct exploitation is by providing specific services based on the result; for example, providing specialised consultancy or contract research. Using the results in standardisation activities is another form of direct exploitation. Partners who are policymakers might also choose to use the results themselves in creating new policy measures.
- *Indirectly:* partners facilitate use to third parties. This could be achieved through the transfer or licensing of results. It should be noted that the creation of a spin-off is thus always considered to be an indirect exploitation of the results. The EC

has recently set up a dedicated service that is focusing on indirect exploitation: the Horizon Results Platform[2].

The methodology for identifying the KERs of the SUN project was a structured, multi-stage process:

1. *Initial Identification of Key Innovation Results (KIRs)*. The consortium first participated in the EC's Innovation Radar initiative. This established methodology provided a bottom-up analysis of the project's innovations, assessing their potential and maturity. This process yielded an initial, comprehensive list of KIRs.
2. *Initial KER Identification*. Following the Innovation Radar analysis, but while research activities were still ongoing, a formal KER survey was distributed to all consortium partners. This survey required partners to identify from the list of KIRs they contributed to, if any of these met the requirements of a KER (i.e., it responds to the specific needs of a well-defined target group, it solves these needs better than the state of the art, and it is chosen by the partner for direct or indirect use). Follow-up discussions with partners helped consolidate the initial list of identified KERs.
3. *Final KER Identification*. After the main research activities were completed and the components of the SUN integrated platform had their final release, a final KER survey was distributed to the partners. At this stage, partners had a much better understanding of the final results than during the initial KER survey. Moreover, an exploitation workshop discussed with partners at large the differences between KIRs and KERs, and when they can consider a result to be a KER. A special focus has been put on the identification of joint exploitation opportunities, where several partners have contributed towards a KER. As a result of the survey, new KERs were identified while previous ones were updated. This survey and the followup discussions that ensued allowed the consortium to validate the final list of KERs.

23.2.3 Steps for Creating Exploitation Plans

With the KER list established, we proceeded to create tailored exploitation plans based on each KER's maturity and potential. The exploitation manager, IN2, organised business clinics in the form of online 1-to-1 exploitation workshops with KER owners. All exploitation workshops had a common methodology (described below), and depending on the discussion with the KER owner, some aspects were delved into more detail than others.

[2] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/horizon-results-platform

- Introduction to the purpose of the session and the methodology used;
- Discussing the main characteristics of the KER and what the possible business models could be considered. This facilitates the follow-up discussion on the specific Unique Value Proposition (UVP) and the creation of a Business Model Canvas;
- Exploring the stakeholder map, with a focus on macro-environment, and identifying relevant stakeholders from different categories (political, economic, social, technological, legal, environmental);
- Developing the UVP using the Value Proposition Canvas method [Osterwalder et al. 2015]. This approach helps by providing a customer-centric view on the benefits brought by the KER, by helping understand the customer needs and challenges, and forcing the participants of the workshop to think from the perspective of the customer. By mapping the customer's "jobs-to-be-done," "pains," and "gains" against the KER's "pain relievers" and "gain creators," we ensured a strong and clearly defined product-market fit. Developing a Unique Value Proposition (sometimes referred to also the Unique Selling Proposition or USP) is a key outcome for this initial exploitation planning because it helps differentiate the offering from competitors, clearly articulating the distinct benefits and value that make the product attractive to target customers and can provide guidance to the KER owner also in terms of further directions of development and research;
- Using the Business Model Canvas [Osterwalder and Pigneur 2010] in order to elicit the key elements of the KER's exploitation plan. The Business Model Canvas is used to provide a comprehensive and easy-to-understand overview of a company's strategic details, including its offerings, infrastructure, customers, and finances. It facilitates structured discussions around management and alignment of business activities, making it an essential tool for capturing, discussing, and iterating on key business elements in a visual format.

Additionally, the project received support from the Horizon BOOSTER programme, an EC-funded service which is helping EC projects navigate the complexities of dissemination and exploitation by providing consulting-like expert support. The SUN project applied to be part of this programme and was accepted for the "Go-to-Market Support Services", which, as its name suggests, helps projects focus on increasing the maturity level of their KERs, developing effective business plans to facilitate market access and impact. The BOOSTER programme provided for this a clear methodology for achieving this, starting from better defining the exploitation intention and filling in the market definition canvas, and continuing with a deep dive into the value proposition canvas

to then draft the exploitation roadmap, carry out a risk assessment and priority map and finally use the Lean Canvas method to define the main elements of the strategy. This process was completed for each of the KERs that were selected to be part of the programme (3 in total). Through the expert guidance and the ensuing discussions and reflections during dedicated workshops, it was possible to arrive to a mature exploitation plan. The combined methodology selected ensured that all project results received bespoke exploitation planning, matching high-intensity external support with the most promising results while guaranteeing a robust baseline strategy for the entire portfolio.

23.3 Market and Business Landscape

The XR market globally is characterised by rapid growth, substantial technological advancements, and take-up across various industries. This trend has been accelerated by the COVID-19 pandemic, where businesses across the board have looked for new digital solutions.

According to [Fortune Business Insights 2024], the XR market was valued at $131.54 billion and is projected to grow to around $183.96 billion by the end of 2024. In the long run, the potential seems huge: $1,706.96 billion by 2032, showing a Compounded Annual Growth Rate (CAGR) of approximately 32.1% during this period. The market estimates from Statista [Statista 6072] are a bit more conservative, pointing out that the XR B2C market size in 2023 was $31.1 billion, with a growth rate of 23.5%; for 2024, the estimated XR Business to Consumer (B2C) market size should be around $38.4 billion.

The market analysis data indicate that although North America has the largest XR market share (due to the presence of multiple large market players), and Asia Pacific is expected to grow with the highest CAGR, Europe is in a good position thanks to the presence of important telecoms companies and an increased investment in innovation based on 5G/6G and XR. Recent efforts spearheaded by the VR/AR Industrial Coalition [Vigkos et al. 2022], which was first announced in 2020 in the Commission's Media and Audiovisual Action Plan and hosts around 200 key European organisations, and the newly established European Partnership on Virtual Worlds demonstrate the strong interest in this sector in Europe from both industry and policymakers. The aim of the initiative and the Partnership is to build sustainable, inclusive, and trustworthy virtual environments across Europe. This will be done through the establishment of a Virtual Worlds Association as a legal entity and creating a Strategic Research and Innovation Agenda (SRIA), which covers several key domains (including those addressed in SUN,

i.e., industry and healthcare) and will help guide the Horizon Europe calls from 2026 onwards.

Gaming and entertainment applications are currently the main drivers of the XR market growth. However, it is expected that Business-to-Business (B2B) solutions will play a much larger role in the next few years. A limiting factor currently is the high implementation costs needed to create credible digital twins for the XR experience.

An important trend observed has been the introduction of smaller and more comfortable XR devices. This is important because an important aspect for the user remains the ease of wearing devices such as headsets, as also revealed by our own research into user requirements for the envisioned SUN pilots. This, coupled with the developments in software applications and connectivity (such as the underlying technological advancements researched in SUN), is expected to support the growing adoption of XR in different sectors and domains. According to [Statista 2021], healthcare (addressed by SUN Pilot 1), manufacturing (addressed by SUN Pilot 2), and the automotive industry are expected to be among the most disrupted by XR technologies. This demonstrates that the SUN approach is perfectly aligned with the most important and impactful market trends.

23.4 Key Exploitable Results

Following the methodology described in the previous section, we present below the KERs of the SUN project.

23.4.1 SUN Integrated Platform

The SUN XR integrated platform (Chapter 16) contains the open SUN platform architecture, as well as various hardware and software components that facilitate the realisation of state-of-the-art XR applications. The platform integrates wearable devices, artificial intelligence, and immersive technologies to make virtual experiences feel more lifelike. It also includes tools to securely manage digital content and protect against cyber threats.

The platform has been used within the SUN project to realise 3 distinct applications, each for a specific vertical scenario: physical rehabilitation, cerebral rehabilitation, and Industry 5.0. These applications demonstrate the versatility and broad technological capabilities of the SUN platform.

The KER is a joint ownership of IN2, which will act as the exploitation lead, ENG, CNR, TG, TUC, HOLO, SSSA, CERTH, UPV, EPFL, and UOG. The platform is envisioned to be open for different types of actors/customers:

- Integrators and consultants, who will provide solutions to customers and are seen as the main target customers;
- XR Content producers, who will create assets to be used in the XR applications;
- Software component creators (e.g., developers and researchers), who maintain existing components or provide new ones;
- Hardware providers (e.g., actuators, headsets), that provide the different hardware devices that are used by the platform.

Indeed, the vision for the SUN Integrated Platform is to become an ecosystem enabling the efficient creation and deployment of high-tech, human-centered XR applications. The exploitation of the SUN platform will take place primarily through service provision, technology transfer, and pilot-based deployments, establishing a pathway toward licensing, professional training, and eventual commercialization. Its modular and multi-component design enables flexible adoption—either as a full platform or through the exploitation of individual KERs—depending on the target market segment.

The integrated SUN platform will be made available under an open core model, combining open architecture with interoperable and reusable components, open hardware, and open-source components with optional proprietary or commercial modules. This hybrid approach allows broad community engagement and integration while supporting business-driven extensions for industrial clients.

23.4.2 Automatic Avatar Production Pipeline

The Automatic Avatar Production pipeline (Chapter 2) enables XR developers to access realistic avatars, on demand, in a scalable, automatic, fast, and cost-efficient manner, without the need for specialised know-how and artistic skills. The avatars are interoperable and reusable in all applications across domains (fashion, healthcare, gaming, wellness, etc.).

The Avatar Production pipeline takes as input a short video produced by the user following simple and clear instructions, and then, in an automatic manner, produces within minutes the Avatar 3D asset, ready to be used (in compatible development engines like Unity). The Avatar is animatable, realistic, and precisely represents the actual user in the very moment they were scanned.

The solution enables a B2B2C chain from XR, Metaverse, and Virtual World developers to vertical businesses in sectors of healthcare, industry, fashion, wellness, and fitness, to final customers in these domains. Avatars with precise representation of real people are powerful tools to develop vertical Use Cases and digital solutions in these markets, enabling a chain of value starting from Avatars.

THING, the SUN partner who owns this KER, plans to exploit the result commercially. The Avatar production pipeline is designed as a Software as a Service (SaaS) asset. Thus, users will access it individually and will be able to produce their avatars through an online and easy-to-use platform. Avatars can be sold as units to individuals for an eventual B2C line, or the platform can be made available with a license scheme for B2B customers, like developers, etc.

23.4.3 MyoLink: A Medium-Density EMG and IMU Interface for Gait Rehabilitation and Movement Analysis

MyoLink allows doctors to get quantitative muscle activation information in real-time and a summary report at the end of rehabilitation sessions, enabling them to provide optimal personalised exercises, feedback, and motivate patients, while requiring only a couple of minutes of set-up time.

MyoLink is a fully integrated, wireless system that combines medium-density Electromyography (EMG) and Inertial Measurement Units (IMUs) to provide real-time, actionable data on muscle activation and kinematics (see Chapter 11). The system consists of elastic bands with embedded electrode arrays and IMUs, allowing for a complete setup in under two minutes without the need for precise electrode placement. The proprietary software, which leverages a deep learning pipeline and a musculoskeletal model, translates complex biosignals into clear, reproducible insights for clinicians and researchers.

This enables personalized, data-driven therapy and objective progress tracking in gait rehabilitation and other movement analysis applications. For instance, the device can be used to "reconnect" people with motor disorders with digital interfaces to reduce isolation in their everyday lives, by allowing them to interact again with digital devices. It can also be used by healthy users to play VR games in an immersive way, or to improve rehabilitation processes by understanding the remaining muscle activity to guide the rehabilitation process. There are also promising results to improve robotic prosthetic hand control.

EPFL, the partner that owns this KER, is currently looking into setting up a spin-off company to bring this result to the market. The initial market entry will be through a paid Pilot Program with selected research laboratories and clinical partners. During

this one-year program, participants will receive the non-certified MyoLink system and dedicated technical support from our team. This phase will focus on generating robust clinical validation data and user testimonials, which will be instrumental for subsequent regulatory submissions and full commercial deployment. Following successful validation and regulatory approval (CE mark, FDA clearance), the spin-off will proceed to a full commercial launch, targeting rehabilitation clinics, physiotherapy centers, and hospitals. The business model will be structured around both hardware sales and recurring service revenues, following a "razor-and-blades" strategy: hardware sells, SaaS, consumables, and licensing.

23.4.4 Tokenized Platform

The Tokenized Platform (Chapter 17) exploits the blockchain technology (like NFTs) and the data space concept for offering services and tools able to manage digital assets in a secure and transparent way, enabling the definition of access policies, revenue models, and collaborative mechanisms for the redistribution of content. The platform offers integration with the Data Spaces infrastructure through integrated data space connectors.

SUN partner ENG plans for a B2B business model, with large businesses, or a data space ecosystem implementer as the main customer. The end-customers are content creators, which could be monetized, for example, by charging a fee on sales, and content consumers/buyers. Service or software providers, such as tech SMEs implementing AI and in need of raw data, could be additional customers.

In addition to the commercial exploitation plans described above, ENG envisions further use of the Tokenized Platform also in other R&D and Innovation projects. ENG is involved in several Data Space deployment projects and will use this know-how and network for further exploitation.

23.4.5 XR Threat Detection

XR Treat detection (Chapter 18) reduces the risk of reputation damage and increases the value of XR products by detecting and stopping, also in real time, several security threats.

It is a backend component working in the background of an XR experience (as one more app on the device), monitoring characteristics and operating data of the device, which are indicative of a security compromise. When it detects a possible compromise, a warning is displayed to the user so that an immediate action can be triggered either manually by the user (e.g., disconnect) or by the application.

UoG, the SUN partner that owns the KER, is interested in commercial exploitation, possibly by creating a spin-off. To accelerate the go-to-market, UoG applied to the Cyber ASAP accelerator programme, which is supported by the UK government. The business model identified is that of a solution as a capability inside the device, with the OS/device manufacturer as the customer. In the future, though, it is envisioned that the business model could be centered around the device owner, as they would purchase and install the solution as an app on their devices.

23.4.6 Hololight Stream

Hololight Stream (Chapter 4) unlocks limitless XR performance by streaming high-fidelity, real-time visuals from powerful servers to any OpenXR-compatible device, bypassing hardware limits. With enterprise-grade security and seamless integration into leading engines like Unity, Unreal, VRED, and Omniverse, it delivers uncompromised, secure 3D experiences anywhere.

The Hololight Stream is an existing product of partner HOLO, and background IPR was brought in at the start of the project. During the project, specific improvements and new functionalities were developed, such as audio support with microphone input and audio extension to other devices, spatial anchor support, camera stream from the XR mobile device to the server, latency reduction, etc.

The results from the project are already being integrated into the commercial product Hololight Stream and offered as a SaaS product with a subscription for the SDK license and additional enterprise plans.

23.4.7 Wearable Thermal Feedback Device for Somatosensory-impaired Users

The wearable thermal feedback device (Chapter 6) is able to restore sensations of touch and caress and provide a more immersive XR experience through accurate and localised thermal stimulation.

The device is composed of an element that is placed on the skin and can change the temperature locally (cool down or heat), thus providing haptic sensations. This interaction is important for both healthy people and patients that had a stroke, amputation, or multiple sclerosis.

Since such similar devices do not exist off the shelf, the KER has significant innovation potential. A related patent application has been filed by EPFL, the SUN partner that owns this KER. A further use of the result is planned as part of upcoming research activities.

23.4.8 Automated Posture Assessment

The Automatic Posture Assessment (Chapter 12) can enable accurate postural assessment from anywhere, effectively enabling remote monitoring and assessment of rehabilitation exercises. The solution can be easily integrated with other systems or in XR applications for rehabilitation.

Other solutions that aim to provide a similar service often use only RGB cameras. While this approach is much cheaper, it suffers significantly in terms of accuracy, which is a key aspect for healthcare professionals.

Currently, the solution is focused on the lower limb (knee pathologies) and works with a list of predefined exercises. The software, which is the main aspect of this KER, uses off-the-shelf hardware (sensors).

The underlying machine learning model that powers the automated posture assessment is planned to be released as open source. THING, the partner that owns this KER, is planning to continue the close collaboration with partners TUC and CERTH for further developments, especially for the use of the results in further research projects and scientific work. THING is also planning to include this KER in its product portfolio, offering it as a B2B solution to large rehabilitation centers, including hospitals, large gyms, and large (research) institutions that work on rehabilitation and biomechanical engineering.

23.4.9 Fingertip Haptic Devices

The developed Fingertip Haptic Devices (Chapter 8) are mature prototypes that can provide advanced, modulated haptic feedback while being compliant with modern VR devices and can thus become a solid basis for a new wearable product line.

Providing tactile feedback in VR is a relevant aspect for effective user interaction. On the other hand, at the moment most of the haptic interfaces tend to be cumbersome (i.e., soft exoskeleton gloves, multi-degree-of-freedom thimbles) or provide scarcely informative feedback (i.e., only vibrotactile, similar to mobile phones vibration). Bulk solutions are also a problem for vision-based tracking systems, now becoming the standard hand tracking method in VR headsets.

Two Fingertip haptic devices have been developed by partners SSSA. They both exploit soft interfaces with the skin to enhance ergonomics and transmission of signals to the fingerpad or to the forearm tissues, and feature a direct-drive, noiseless actuation. Compared to other solutions proposed in the state of the art, they provide low-intensity, but clean, dynamic, and informative haptic signals. The compact shape makes them

compliant with vision-based hand tracking systems featured by commercial 3D visors (i.e., Oculus Quest, HoloLens). SSSA has placed a special focus on features such as wearability and compliance in order to obtain better usability and acceptability of haptic interfaces by end users.

The first device will be released as open source, while the method used in the second device is currently part of a filed patent (pending). Thus, there are three different exploitation paths that have been identified: a) use in further research (either own research or paid by a company) based on the open source solution; b) knowledge transfer or licensing to a company doing wearable devices; and c) patent royalties.

23.4.10 OmniBridge

OmniBridge (Chapter 16) is a middleware that aggregates, standardizes, secures, and manages data flow within the SUN ecosystem. It is a flexible, decentralised, and performant data broker implemented on top of Java Quarkus, MQTT, gRPC, and protocol buffers. By providing secure APIs and message-based communication, OmniBridge enables other technology providers to integrate new modules into the SUN platform with minimal effort — only a thin adaptation layer is needed on top of their existing components.

While OmniBridge has been tailored to the architectural needs of the SUN project, its capabilities extend far beyond. It represents a generic data-exchange backbone that can support any distributed or hybrid platform requiring efficient communication among heterogeneous components. Its differentiating strengths lie in its real-time performance, security-by-design architecture, and scalability across on-premises and cloud deployments.

Partner IN2 plans to exploit OmniBridge through a dual strategy combining research and commercial initiatives. In the research domain, the middleware will serve as a foundation for new Horizon Europe and national innovation projects that demand flexible integration frameworks for AI-driven or immersive systems. On the commercial side, IN2 will incorporate OmniBridge into its portfolio as part of custom B2B data-exchange and interoperability services, enabling clients to deploy secure, high-throughput backends tailored to their operational environments. This approach not only broadens IN2's service offering but also positions OmniBridge as a core enabler of next-generation, interoperable digital ecosystems.

23.4.11 Dense Vision-Language Models for Open-Vocabulary Understanding

Novel multimodal models developed in SUN by partner CNR are capable of efficiently bridging vision and language in a dense manner, i.e., describing the local semantics of an image or a 3D object/scene for every one of its pixel/location in a very efficient way (see Chapter 5). This expands the tasks that Large Multimodal Models can solve in generic domains (i.e., in open-vocabulary or zero-shot settings).

Attaching dense local semantics to images opens up new, simplified, and efficient methodologies to tackle visually-local tasks with natural language. For example, they could enable applications like a) counting small objects in an image described by a textual query, b) provide a textual description of a local area of an image (e.g., visually impaired people could get a description of the area of an image that's under their finger while swiping), and c) searching for sub-regions in image collections with textual queries, all in a training-free and zero-shot setting. The results have been openly published [Bianchi et al. 2024; D'Orsi et al. 2025; Bianchi et al. 2025], and CNR is planning to use these in future research activities.

23.4.12 Advanced AI-driven 3D Design, Reconstruction, Processing, and Optimization

CNR developed a suite of methods that enhance various stages of the 3D processing pipeline: from 3D asset acquisition and reconstruction [Callieri et al. 2025], to texture enhancement [Maggiordomo et al. 2023], and shape design and optimization [Favilli et al. 2024] (see Chapter 1).

These research outcomes significantly reduce the time and effort required to produce high-quality, usable assets or to improve existing ones in terms of quality and performance. They also have the potential to open up new research directions and application opportunities. The results have been openly published, and CNR is planning to use these in future research activities.

23.4.13 Policy Paper on Ethics

During the SUN project, partner VUB conducted benchmark research on the legal and ethical framework for XR technologies (see Chapter 25). The study established a regulatory and compliance framework for the project and contributed to the broader discussion on ethical and lawful XR development in Europe.

Since no specific EU regulation yet governs XR, the research analysed how existing instruments, such as the GDPR, AI Act, and Data Governance Act, apply to immersive environments. It also identified regulatory gaps and ethical risks related to data collection, privacy, and user safety in XR applications. VUB also developed different templates, e.g., for data protection assessment, which could be reused openly.

The work addressed three key stakeholder groups:

- *Policy makers*, who can use the framework to guide future ethical XR regulation;
- *Industry actors*, seeking practical compliance tools for data protection and consent management;
- *End users*, who need greater awareness and empowerment regarding their digital rights.

The learning from the above has been synthesised in a policy paper which can underpin future research and policy initiatives on responsible XR innovation.

23.5 Conclusions

The SUN project established a comprehensive framework for transforming cutting-edge XR research into sustainable innovation. By combining systematic KER identification, EC-aligned methodologies, and targeted business modeling, the project aims to achieve the real exploitation potential of the results, which range from a modular integrated XR platform to an ethical framework. While the human-centered design and open innovation approaches used in the project are key drivers of societal and economic impact, some challenges remain: many technologies still require further validation, productisation, regulatory adaptation, and large-scale testing before full market adoption. Future research could address the further development of the state-of-the-art technologies, interoperability standards, ethical data governance, and the creation of European value chains supporting XR deployment.

References

APRE and CDTI (2022). *Guiding notes to use the TRL self-assessment tool*. Tech. rep. Version FINAL2. BRIDGE2HE – H2020-101005071. URL: https://horizoneuropencpportal.eu/sites/default/files/2022-12/trl-assessment-tool-guide-final.pdf.

Bianchi, Lorenzo, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi (2024). "The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22520–22529.

Bianchi, Lorenzo, Fabio Carrara, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi (2025). "Fine-grained Open-vocabulary Object Detection". In: *Under Review*. URL: https://doi.org/10.5281/zenodo.17339026.

Callieri, Marco, Massimiliano Corsini, Somnath Dutta, Daniela Giorgi, and Marco Sorrenti (2025). "AI-driven specular removal for 3D asset creation". In: *Proceedings of the 25th International Conference on Digital Signal Processing (DSP 2025), Special Session on Digital Twins and XR: Signal Processing Challenges and Emerging Technologies*. Presented at DSP 2025.

Commission, European (2020). *Horizon results booster – Helps to bring a continual stream of innovation to the market and beyond*. PDF available via Scribd. DOI: doi/10.2777/540837.

D'Orsi, Domenico, Fabio Carrara, Fabrizio Falchi, and Nicola Tonellotto (2025). "Breaking the 2D Dependency: What Limits 3D-Only Open-Vocabulary Scene Understanding". In: *2025 International Conference on Content-Based Multimedia Indexing (CBMI)*. In Press.

Favilli, Andrea, Francesco Laccone, Paolo Cignoni, Luigi Malomo, and Daniela Giorgi (2024). "Geometric deep learning for statics-aware grid shells". In: *Computers & Structures* 292, p. 107238.

Fortune Business Insights (2024). *Extended Reality Market Size, Share & Industry Analysis, By Type (Virtual Reality, Augmented Reality, and Mixed Reality), By Industry (Healthcare, Education, Retail & E-commerce, Gaming, Automotive, Media & Entertainment, and Others), and Regional Forecast, 2025–2032*. Report ID FBI106637. Accessed: 10-05-24. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/extended-reality-market-106637.

Hirzel, Simon, Tim Hettesheimer, Peter Viebahn, and Manfred Fischedick (2018). "Bridging the valley of death: A multi-staged multi-criteria decision support system for evaluating proposals for large-scale energy demonstration projects as public funding opportunities". In: *eceee Industrial Summer Study Proceedings – "Policies and Programmes to Drive Transformation"*, pp. 105–115.

Maggiordomo, Andrea, Paolo Cignoni, and Marco Tarini (2023). "Texture inpainting for photogrammetric models". In: *Computer Graphics Forum* 42.6, e14735.

Osterwalder, Alexander and Yves Pigneur (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons.

Osterwalder, Alexander, Yves Pigneur, Greg Bernarda, Alan Smith, and T. Papadakos (2015). *Value Proposition Design: How to Create Products and Services Customers Want*. John Wiley & Sons.

Statista (2021). *Sectors expected to witness the most disruption by immersive technologies according to XR industry experts worldwide in 2021*. https://www.statista.com/statistics/1185060/sectors-disrupted-immersive-technology-xr-ar-vr-mr/. Statista No. 1185060. Accessed: 10-05-24.

Vigkos, A, D Bevacqua, L Turturro, S Kuehl, T Fox, P Diestre, and S Stig Yding (2022). *The virtual and augmented reality industrial coalition: strategic paper*. Publications Office of the European Union.

24. Impact: A Multi-Perspective View

Fabio Perossini[1], *Giuseppe Caracciolo*[1], *Silvia Boi*[1], *Devin Bayer*[2], *Michaela Wiese*[2], *and Claudio Vairo*[3]

[1] KPeople Research Foundation (KPRF), Malta
[2] Outdoor Against Cancer (OAC) Germany
[3] Institute of Information Science and Technologies, National Research Council (CNR-ISTI), Italy

Abstract. This Chapter presents the SUN (Social and hUman ceNtered XR) project's strategy for linking technological innovation with human values in Extended Reality (XR). It highlights how AI, immersive interfaces, and multimodal interaction were piloted in two key domains: clinical neurorehabilitation and Industry 5.0. In healthcare, adaptive XR environments and wearable-based interactions improved patient motivation, adherence, and recovery, while reducing caregiver burden. In industrial contexts, digital twins, haptics, and AI-driven learning enhanced worker safety, inclusivity, and training efficiency. The Chapter emphasizes co-creation with patients, clinicians, workers, and safety experts to ensure usability, trust, and acceptance. Impact monitoring extended beyond technical KPIs to include ethical and experiential dimensions such as fairness, transparency, and user satisfaction. Results show both short-term gains—improved rehabilitation outcomes and workplace safety—and long-term pathways, including integration into European Data Spaces and refinement of ethics-by-design guidelines. The study concludes that SUN represents not just a technological advance but a model of human-centred, responsible innovation with strong clinical, industrial, and societal relevance.

24.1 Introduction

The SUN (Social and hUman ceNtered XR) project explores how Extended Reality (XR) technologies, when combined with Artificial Intelligence (AI) and multimodal interaction, can generate meaningful societal and industrial impact while remaining deeply rooted in human values. Recognizing that immersive technologies represent not only technical advances but also socio-cultural shifts, SUN adopts a framework where innovation is guided by inclusivity, ethics, and real-world applicability. This Chapter presents a multi-perspective view of impact, reflecting on how the project's pilots—in clinical neurorehabilitation and Industry 5.0 workplaces—were conceived, implemented, and evaluated. The emphasis is on participatory design: scenarios were co-created with therapists, patients, workers, and safety experts to ensure relevance, usability, and acceptance. The approach highlights the project's commitment to human-centricity, where impact is measured not only by technical performance but also by qualitative factors such as trust, fairness, transparency, and user well-being. By coupling technological piloting with continuous human-centred monitoring, SUN demonstrates that XR innovation can contribute simultaneously to clinical recovery, workplace safety, and broader societal objectives. The next paragraphs set the stage for examining the project's methodology, results, and the anticipated medium- and long-term trajectories of its innovations.

24.2 SUN: Coupling Humanity and Technology

The SUN project presents a forward-looking framework that emphasizes the symbiosis between human values and technological advancements in the realm of XR. With the dual ambition of advancing clinical rehabilitation and supporting Industry 5.0 scenarios, SUN integrates immersive XR technologies with AI-based methodologies through a deeply human-centric lens. The pilot definition process, monitoring methodologies, and impact assessment approach are all underpinned by a philosophy that centers human needs, capacities, and aspirations.

24.2.1 Human-Centric Piloting

The foundation of the SUN project lies in the participatory definition of pilot scenarios. These pilots were designed for two primary domains: clinical neurorehabilitation and human-centric industrial workplaces. Through stakeholder engagement and co-creation with therapists, patients, industrial workers, and safety experts, the scenarios were articulated around real-world needs, opportunities, and constraints.

In the neurorehabilitation domain, the scenarios build upon empirical evidence indicating the value of XR and AI in enhancing motor recovery, neuroplasticity, and patient motivation during the rehabilitation process. Here, SUN developed adaptive exergaming environments and wearable-based interaction systems tailored for home use and clinical settings alike, addressing specific requirements of Acquired Brain Injury (ABI) patients, such as safety, usability, feedback, and personalization.

In the context of Industry 5.0, SUN framed XR scenarios for safe and inclusive worker training, collaboration, and ergonomics enhancement. Digital twins, AI-enhanced learning loops, and multisensory haptics were used to ensure that virtual environments could be adapted to the individual capacities, stress levels, and preferences of the workforce.

24.2.2 Technology perspective Piloting

The SUN project adopted a modular and user-oriented technology piloting strategy that emphasized the seamless integration of AI, XR, and multimodal interaction systems. Leveraging a multi-layer platform architecture, SUN integrated a wide range of components, such as wearable haptic devices, AR/VR interfaces, pose estimation systems, AI-based emotion and gesture recognition, personalized avatars, and cyber-threat monitoring modules, through a unifying middleware to ensure system interoperability and responsiveness. This approach enabled the project to support diverse use cases through a shared technological backbone, adaptable to individual user needs and contextual constraints. Real-time interaction, personalization, and dynamic adaptation were core technological pillars, ensuring that system behavior could respond meaningfully to user input, workload, and emotional state.

Across both clinical and industrial domains, pilots served as testbeds for validating the feasibility of running intelligent XR solutions across different environments using lightweight, wearable-friendly hardware. The emphasis was not only on functional performance but also on usability, latency, integration of feedback modalities, and edge-computing compatibility. SUN's piloting methodology thus provided robust validation of its technological vision while reinforcing its commitment to scalable, inclusive, and ethically aligned innovation.

24.2.3 Continuous Human-Centred Monitoring of Innovation

Beyond scenario design, the SUN approach incorporates a framework for continuous innovation monitoring that aligns with human-centric principles. Rather than merely measuring technical Key Performance Indicators (KPIs), the framework includes qualitative and mixed-method evaluations focusing on human experience (e.g., Quality of

User Experience (QUX)), social inclusion, and Ethical-Legal-Social Implications (ELSI). This approach draws from the HINTS (Health Information National Trends Survey) metrology framework, combining formal concept analysis with a multi-dimensional ethical and data privacy model. Innovations were tracked according to their performance on attributes such as fairness, transparency, respect, and interpretability. These assessments were embedded within the life-cycle of each pilot to support adaptive development and early detection of unintended consequences.

The SUN project culminated in an impact evaluation that accounted for both immediate outcomes and long-term expectations. For clinical rehabilitation, the project demonstrated improvements in patient engagement, recovery adherence, and reduced caregiver burden. For industrial workers, XR tools enhanced procedural safety, accelerated training timelines, and increased satisfaction with digital interfaces.

Importantly, the SUN methodology enables the anticipation of medium-term impacts (three years post-project), leveraging digital twins and federated data spaces to sustain adaptive learning across applications. Planned post-project actions include the refinement of XR ethics-by-design guidelines, large-scale integration into European Health and Industry Data Spaces, and the evolution of personalized avatars as dynamic digital identities for users in training and rehabilitation contexts.

24.3 Mitigating Acceptance Through User Experience

The definition and implementation of pilots in the SUN project followed a deliberate and methodologically rigorous path rooted in human-centred design principles. By adopting a participatory approach from the outset, the project ensured that the technology being developed would not only address real-world needs but also resonate with the users' values, capabilities, and contexts. This alignment is vital in XR and AI-driven systems, where complexity and unfamiliarity can pose barriers to adoption.

24.3.1 Grounding Scenarios in the Reality of Users

Scenarios within SUN were not defined in a vacuum. Instead, they emerged from a rich process of co-creation involving clinicians, patients, caregivers, industrial workers, safety managers, and digital designers. This inclusive model helped bridge the gap between abstract technological potential and everyday user experience. In clinical rehabilitation, for example, XR-based exergames were tailored to reflect the reality of patients recovering from acquired brain injuries. Insights from interviews and thematic analyses underscored needs related to safety, feedback, adaptability, and motivational

support, which directly shaped the features of the rehabilitation scenarios. In industrial environments, pilot scenarios focused on fostering safer, more inclusive, and more efficient collaboration through XR-enhanced simulations and digital twin models. Workers were involved early in identifying friction points in training and operations, including ergonomic discomfort, attention overload, and knowledge transfer gaps. These insights shaped the development of haptic feedback interfaces and AI-supported XR instruction modules that could adapt to individual users' pace, stress level, and physical constraints.

24.3.2 Acceptance as a Design Priority

Technological acceptance is often hindered by a disconnect between system design and the socio-emotional reality of its users. SUN addressed this by positioning acceptance not as a post-deployment challenge but as a foundational design criterion. The project drew from experience assessment methodologies such as the QUX model, integrating both subjective perceptions and objective interaction data to continuously inform iterative improvements. Furthermore, acceptance was linked to ethical and social trust; participants' concerns about privacy, equity, and cognitive overload were not dismissed but incorporated into the scenario refinement process. For example, in rehabilitation use cases, the presence of personalized avatars and transparent AI-driven feedback loops were introduced to foster a sense of familiarity, safety, and control.

24.3.3 From Familiarization to Empowerment

By embedding users deeply into the scenario definition and evaluation process, the SUN project promoted not only acceptance but ownership. In clinical settings, this improved therapy adherence and eased caregiver burden. Patients who saw their own data reflected meaningfully in the XR environments—for example, through progress visualizations or adaptive task difficulty—reported higher engagement and trust in the system. In industrial scenarios, the ability to rehearse tasks in a safe, virtual replica of the workplace boosted both competence and confidence. Workers noted that the systems "spoke their language," not just in interface design but in their capacity to adapt to context and respond to individual user profiles. This reflects a deeper shift from passive interaction to active co-evolution of technology and human capability.

24.3.4 How technologies face acceptance

From its inception, the SUN project positioned user acceptance as a core design driver, integrating it into every stage of technological development and validation. Rather than treating acceptance as a post-deployment outcome, SUN approached it as a

continuous, co-evolutionary process, shaped through iterative piloting, multi-sensorial interaction, and responsive system behaviors. Across pilots, SUN adopted multimodal user interfaces, from gaze and Electromyography (EMG)-based control schemes to haptic and avatar-mediated feedback, to ensure inclusivity and adaptability to diverse user capabilities. Acceptance was monitored through structured questionnaires, informal feedback sessions, and real-time interaction analytics, focusing on indicators such as perceived effort, ease of use, safety, and emotional comfort.

In clinical rehabilitation, acceptance was enhanced by making therapy engaging and goal-driven, incorporating gamified environments, transparent feedback on exercise performance, and personalization based on physiological and emotional inputs. In industrial settings, technology acceptance was fostered by aligning interfaces with existing workflows, supporting explainability in AR alerts, and minimizing the intrusiveness of wearable components. The SUN methodology ensured that acceptance was not a one-size-fits-all target, but a dynamic interaction between user needs, environmental constraints, and technological affordances, ultimately enabling a gradual transition from familiarization to empowerment.

24.3.5 The users' viewpoint

The involvement of the Outdoor Against Cancer (OAC) organization underlines the practical and ethical value of integrating organisations that closely work with cancer survivors and patients. While a lot of the work that has been done as part of the SUN project was of a technical nature, bridging technology with a human-focused approach is key, especially when working with vulnerable populations like patients in Pilot 1. Building on its commitment to human dignity, inclusivity, and autonomy (as outlined in Chapter 19), OAC sees the importance of translating ethical principles and the potential of technology into real-world and patient-centered impact, to maximise the positive effects that can be achieved with the development of the SUN platform.

Building Bridges Between Innovation and Everyday Care

While the theoretical groundwork for the SUN project was of utmost importance, the question is always how the potential translates into real-world scenarios and what sorts of applications are possible. For Pilot 1, which contains the use case of Upper Limb Rehabilitation (see Chapter 20), OAC, as representative for patients and healthcare end-users' needs, helped establish conceptual guidelines and acquire cancer survivors for the pilot validation. OAC's long-standing experience in providing training, education, and overall support for cancer survivors and their social environment has built a strong level of trust in the services offered, which significantly helped the process of finding

study prospects that were interested and felt comfortable enough to participate in the pilot of the project. Knowing that patients' well-being is the highest priority can make the pilot and testing experiences to appear easier and safer for the users. The involvement of an organisation directly supporting participants during the testing period created a reliable feedback channel between users and developers. This feeling of trust and safety encouraged more open and honest feedback from participants. By involving a patient-focused organisation like OAC as a consultant and mediator between users, developers, and clinicians, the project has been able to strengthen continuity between clinical goals, technical feasibility, and a positive patient experience. Patient feedback and stakeholder dialogue were essential in ensuring that the SUN platform was not only technically sound but also socially relevant.

A Scalable Model for Human-Centric XR Deployment

There is often an overlooked space between high-tech innovation and on-the-ground therapeutic reality that needs bridging, maneuvering, and open communication. It is essential to ensure that developed tools align with both clinical goals and human needs, translating ethical concerns into actionable design insights, amplifying marginalized perspectives, and fostering trust and informed adoption.

OAC's role in the SUN project serves as an example and replicable model for involving civil society and non-government organizations in the development of XR applications or, even more broadly, in technological development itself. It must be ensured that technological progress does not bypass human needs. As XR, AI, and technology in general continue to evolve, this form of structured, yet multi-dimensional involvement can and should become standard practice in European innovation projects. It reflects core EU values of inclusion, participation, and responsible research, while also contributing to better technology, better outcomes, and a stronger connection between science and society. An approach grounded in co-creation and shared responsibility is essential, not only to create technically advanced tools but to achieve innovation that is meaningful, impactful, and deeply human. Innovation that serves, technology that augments, care that stays unmistakably human.

24.4 Piloting as a Means to Measure Innovation

In the SUN project, pilots were not only instruments for user engagement and co-creation, but also powerful frameworks for evaluating the pace, quality, and direction of innovation. These scenarios allowed the teams to observe how new technologies interacted with human behavior, institutional processes, and social environments over

time. By embedding assessment mechanisms within the real-life deployment of XR solutions, the project transformed each pilot into a live laboratory of innovation.

24.4.1 Scenarios as Living Testbeds

The dynamic and iterative nature of SUN's pilot scenarios turned them into "living testbeds" where technological and social innovation could be monitored in context. This approach allowed the team to move beyond static evaluations of system performance and into multidimensional assessments, including social, ethical, and cognitive impacts. In the rehabilitation domain, metrics have traditionally focused on clinical outcomes such as range of motion or frequency of exercise. In SUN, these were complemented by measures of user engagement, emotional response, and adaptability of the system to evolving user needs. The presence of intelligent avatars, adaptive difficulty mechanisms, and multisensory feedback loops made it possible to trace how personalization algorithms contributed to innovation through increased compliance and reduced frustration over time. For the industrial scenarios, innovation was tracked through changes in learning efficiency, error rates, user feedback, and comfort perception. Workers engaged in XR-based procedural simulations provided insights not only about task completion but also about the intelligibility of interfaces, perceived mental workload, and the sense of control over the learning process. Here, innovation was measured not just by system capabilities but by how well the systems enhanced human capabilities under real operational conditions.

24.4.2 Multidimensional Monitoring: Beyond KPIs

SUN's innovation monitoring framework departed from conventional KPIs by integrating human-centric indicators derived from the ethical and data privacy metrology model developed by the HINTS project. This model employed Formal Concept Analysis to classify innovations not only by technical advancement but also by adherence to attributes such as fairness, explainability, user empowerment, and social relevance. These dimensions were particularly relevant for identifying unintended consequences early. For instance, in the industrial pilot, real-time adaptive guidance systems led to improved task performance but also revealed a temporary loss of perceived autonomy among novice users. By capturing this friction through qualitative feedback and emotional data analysis, the project was able to refine the system to better support both user autonomy and accuracy.

24.4.3 Co-evolution of Technology and Practice

A significant insight from applying scenarios to measure innovation was the observable co-evolution of technology and practice. As users grew more familiar and confident with XR tools, they started requesting new features, challenging early limits, and adapting workflows to exploit XR affordances. This mutual shaping is a clear marker of meaningful innovation, showing that technology is being internalized, adopted, and reshaped by users. This was especially evident in the rehabilitation domain, where patients moved from passive recipients of therapy to active partners in customizing their XR experiences. Through biofeedback loops and personal data visualization, users were empowered to understand and influence their own progress, thereby driving innovation from within the system rather than waiting for external updates.

24.4.4 The Value of Scenario-Driven Metrics

What SUN ultimately demonstrated is that scenario-driven metrics serve a dual function: they validate technological progress while simultaneously revealing new user needs and societal impacts. This recursive quality makes scenario-based measurement especially suitable for complex sociotechnical systems like XR and AI. It allows innovation to be framed not just as invention, but as transformation, a shift in how people live, work, and recover. As XR systems mature and scale, the SUN approach highlights the importance of treating scenarios not as static demonstrations but as evolving ecologies. Only through continuous, situated engagement with users can innovation be accurately measured, ethically guided, and meaningfully sustained.

24.5 Innovation Monitoring and Stakeholders' Collaboration

In the SUN project, innovation monitoring and stakeholder collaboration were intrinsically interwoven processes. The innovation journey did not follow a linear, top-down logic but was instead shaped by iterative, dialogic exchanges between researchers, developers, end-users, and ethical advisors. This collective intelligence approach ensured that technology development remained ethically sound, user-informed, and societally relevant across all stages of the project.

24.5.1 Monitoring Innovation with a Human-Centric Lens

Innovation within SUN was not solely about deploying new technologies; it was about continuously ensuring that these technologies responded meaningfully to the evolv-

ing needs of users. This required an innovation monitoring system that extended beyond conventional benchmarks. SUN developed and adopted a multi-criteria assessment framework based on human-centered innovation indicators, incorporating emotional, ethical, ergonomic, and usability dimensions alongside traditional technical performance. At the core of this framework was a structured yet flexible monitoring loop that included real-time user feedback, behavior analytics, and co-assessment interviews. These mechanisms enabled the SUN team to detect early signs of friction or user discomfort and to adapt system features accordingly. By embedding the capacity for reflexivity, SUN's monitoring practice helped bridge the gap between exploratory research and accountable innovation. One illustrative example is the refinement of haptic feedback tools in industrial XR training. Initial trials revealed that while users appreciated the realism, prolonged use induced fatigue. Continuous feedback loops helped recalibrate the balance between fidelity and comfort, demonstrating how monitoring acted as a steering mechanism for innovation quality.

Stakeholders as Co-Innovators

SUN recognized that stakeholders are not just recipients of innovation but co-creators. Across its pilot domains, the project implemented structured stakeholder engagement activities including co-design workshops, scenario walkthroughs, and ethics consultation panels. These processes aimed to empower diverse user groups, including patients, clinicians, factory workers, designers, and legal experts. In the clinical context, patients with acquired brain injury and their therapists were invited to shape the progression of XR-based exergames. Their involvement influenced not only interface aesthetics and interaction mechanics but also the types of motivational feedback integrated into the system. This deep collaboration contributed directly to the system's relevance and sustainability over time. In industrial contexts, collaboration extended to unions and workplace safety bodies, ensuring that XR tools did not inadvertently introduce cognitive or physical burdens. This dialogical openness fostered mutual trust and informed risk assessments that went beyond regulatory checklists to encompass lived experiences of the workforce.

24.5.2 Institutional Learning and Ecosystem Building

An often-overlooked benefit of continuous stakeholder collaboration is the institutional learning that emerges. In SUN, partners developed shared vocabularies, risk awareness, and innovation literacy that persisted beyond the individual pilots. For example, the ethical metrology framework co-developed with stakeholders became a reusable toolkit applicable to future XR and AI implementations, embedding a culture of anticipatory governance within participating institutions. Moreover, SUN's emphasis on federated

collaboration laid the foundation for a broader ecosystem vision. Stakeholders not only co-designed solutions but also explored post-project pathways for scaling, including integration into national health platforms and cross-sector training programs. This long-term alignment between innovation monitoring and ecosystem thinking underlines SUN's commitment to legacy and continuity.

24.6 The Use of Specific Tools to Capitalise Innovation

The SUN project adopted a structured and participatory approach to monitor and capitalize on innovation, grounded in methodologies that integrated both internal coordination and external engagement. Innovation management was framed as a continuous, reflexive process where new ideas, technological advancements, and ethical implications were jointly evaluated by stakeholders. This ensured that innovation outcomes were not only technically robust but also societally relevant and ethically aligned.

24.6.1 Innovation Monitoring Methodology

The innovation monitoring strategy in SUN was designed to support early detection, assessment, and reuse of valuable innovation assets. The process included five iterative stages:

- *Innovation identification*: Partners registered emerging innovations derived from technical developments, pilot activities, or stakeholder feedback;
- *Eligibility assessment*: Innovations were evaluated based on novelty, relevance, feasibility, and alignment with project goals;
- *Documentation and classification*: Eligible innovations were described using a shared metadata template that captured their technical scope, application potential, and ethical dimensions;
- *Review and validation*: Innovations underwent review processes involving both internal experts and external feedback to ensure robustness and usability;
- *Capitalization planning*: Innovations were selected for further integration into pilot developments, communication materials, or exploitation pathways.

This methodology fostered collaborative innovation tracking and was supported by agile project tools to maintain transparency and accountability.

24.6.2 Stakeholder Engagement for Innovation Co-Creation

Stakeholder engagement played a central role in ensuring that innovation was grounded in real-world needs and expectations. The engagement process was designed to be:

- *Iterative*: Stakeholders were involved throughout all phases of innovation: idea generation, prototyping, testing, and evaluation;
- *Inclusive*: A broad spectrum of stakeholders (including researchers, developers, patients, clinicians, industry representatives, and ethics experts) participated in structured dialogues;
- *Reflective*: Feedback loops ensured that insights from end-users and domain experts were directly translated into design changes and innovation decisions.

Methods included co-design sessions, structured interviews, online surveys, and participatory evaluation workshops. All outputs were systematically analyzed and fed into the innovation monitoring workflow.

24.6.3 Use of Confluence and Jira to Structure Innovation Flow

To coordinate the innovation pipeline and stakeholder contributions, SUN employed dedicated project management tools:

- *Confluence*[1] functioned as a collaborative knowledge space, hosting shared documents, innovation logs, methodological guidelines, and annotated records of stakeholder engagements;
- *Jira*[2] was used to track the status of each innovation asset through its lifecycle. It enabled issue creation, assignment, prioritization, and timeline management for innovations under consideration.

Together, these tools supported traceability, real-time coordination, and integration of insights across teams and use cases.

24.6.4 Capitalization Through Structured Project Outputs

Innovation outputs were consolidated and disseminated through structured documentation and formal project reporting. Each validated innovation was accompanied by: i)

[1] https://www.atlassian.com/it/software/confluence
[2] https://www.atlassian.com/it/software/jira

a detailed technical description and user value proposition; ii) usability and ethical assessments; iii) stakeholder feedback summaries; iv) recommendations for deployment to other domains. These structured outputs ensured that innovations were accessible, interpretable, and positioned for future exploitation beyond the project's duration.

24.7 Human-Centred Impact Assessment

The SUN project placed human experience at the core of its mission, not only in terms of design and development but also in its approach to assessing impact. From the outset, the project articulated a vision of transformative effects across clinical rehabilitation, industrial productivity, and societal acceptance of XR-AI technologies. This vision, outlined in the expected impacts of the project's planning documents, was pursued through a lifecycle-oriented approach to evaluation, measuring both the immediate benefits within pilot activities and the anticipated longer-term transformations enabled by the SUN methodology.

24.7.1 A Lifecycle-Based Approach to Impact Measurement

SUN's impact assessment evolved alongside the project lifecycle. Early stages focused on potential and formative impacts, such as stakeholder awareness, capacity building, and the development of inclusive technological narratives. As pilots matured, the focus shifted to measurable functional, emotional, and ethical outcomes observed in real deployment contexts. Impact measurement was framed across three complementary levels:

- *Micro-level*: User experience, behavioral response, and personal outcomes;
- *Meso-level*: Organizational processes, quality of service delivery, and stakeholder collaboration;
- *Macro-level*: Systemic readiness, regulatory alignment, and public value.

This multilevel approach allowed SUN to track not just what changed, but how and for whom it changed.

Impact in Clinical Rehabilitation

SUN demonstrated meaningful progress toward improving clinical rehabilitation outcomes by enabling personalized, adaptive, and motivating XR-based experiences for patients recovering from acquired brain injuries. Key impact points included:

- *Enhanced adherence to therapy*: Patients reported higher motivation and engagement due to interactive and gamified rehabilitation interfaces;
- *Empowerment and self-awareness*: The use of personalized avatars and visual progress indicators improved patient understanding of their recovery paths;
- *Therapist support and workload optimization*: XR systems offered tools for remote monitoring and adaptive feedback, reducing pressure on clinical staff while preserving the quality of care.

These effects directly addressed the project's impact goal of reducing healthcare burden and supporting decentralized rehabilitation pathways.

Impact in Human-Centric Industry 5.0 Environments

In industrial settings, SUN introduced human-centred XR applications that enhanced learning, safety, and inclusion. Through its pilots, the project showed:

- *Faster upskilling and higher retention*: Workers undergoing XR-based training completed onboarding tasks more quickly and with fewer errors;
- *Inclusive design*: The integration of adaptive features, such as task pacing and environmental adjustment, supported diverse worker profiles, including those with limited prior exposure to digital technologies;
- *Reduction of cognitive and physical strain*: Ergonomically validated XR tools helped prevent fatigue and improved perceived task comfort.

This progress aligned with broader Industry 5.0 goals to reinforce human-machine collaboration while enhancing well-being and workplace equity.

Ethical, Social, and Policy-Oriented Impact

SUN's emphasis on trustworthy and transparent innovation translated into structured ethical and social impacts:

- *Increased trust in AI systems*: Transparent decision support, explainable feedback mechanisms, and privacy-by-design architectures promoted stakeholder confidence;
- *Improved stakeholder literacy*: Regular engagement sessions helped users and decision-makers understand and contribute to technological co-creation;

- *Contribution to standardization and policy reflection*: SUN generated replicable ethical evaluation tools and contributed insights relevant to European AI and XR governance frameworks.

These outcomes reflect the project's intention to go beyond compliance and foster a culture of proactive responsibility in immersive technology development.

24.7.2 Towards Long-Term Impact and Sustainability

SUN's work was not limited to immediate pilot outcomes. The project also initiated processes expected to yield impact over the next 3–5 years. The assessment of long-term impact was structured around two key dimensions: scale and significance.

Scale: The scale of SUN's expected long-term impact is reflected in the replicability and transferability of its technological solutions and methodological frameworks. Core outputs such as the modular XR-AI platform, human-centric digital twins, and ethical innovation monitoring methodology were designed to serve diverse sectors and user communities beyond the original pilot contexts. The following dimensions illustrate how the project's results can be scaled and transferred:

- *Cross-sectoral application*: Tools developed in SUN are adaptable to other domains such as mental health, neurodegenerative disease care, vocational training, logistics, and smart manufacturing;
- *European alignment*: The project's architecture is interoperable with emerging European digital infrastructures such as the European Health Data Space and industrial Data Spaces, supporting a continent-wide scaling strategy;
- *Ecosystem integration*: SUN established collaborations with innovation hubs, hospital networks, and industrial clusters, ensuring that results are embedded in live ecosystems capable of sustaining and expanding their use.

Significance: The significance of SUN's long-term impact lies in the depth and sustainability of transformation across stakeholder groups and systems:

- *Empowered patients and workers*: SUN technologies foster self-efficacy, personalized progression, and cognitive-emotional safety, enabling sustained adoption and improved quality of life or work;
- *Organizational change*: Institutions adopting SUN outcomes gain new capabilities in XR-supported service delivery, digital twin analytics, and ethics-aware innovation governance;

- *Policy and standardization influence*: SUN contributes concrete methodological models and technology guidelines that are informing standardization activities and helping shape regulatory best practices.

24.8 Conclusions

The SUN (Social and hUman ceNtered XR) project demonstrates how Extended Reality, when combined with Artificial Intelligence and a rigorous human-centric framework, can generate tangible impact across clinical, industrial, and societal domains. Through participatory pilot design, SUN validated that technology adoption is most effective when grounded in user needs, ethical principles, and continuous feedback loops. In neurorehabilitation, the project showed how adaptive XR environments, multimodal feedback, and personalized avatars can improve patient motivation, therapy adherence, and recovery, while reducing the burden on caregivers. In industrial contexts, XR-enhanced simulations and digital twins supported safer, more inclusive, and efficient workplaces, fostering trust and engagement among workers. Beyond immediate outcomes, the project anticipated medium- and long-term trajectories, such as integration into European Health and Industry Data Spaces, the evolution of digital identities, and the refinement of ethics-by-design guidelines. By placing acceptance and usability at the centre of design rather than treating them as afterthoughts, SUN provided a replicable model of responsible innovation. Its impact extends beyond technical progress, positioning XR as a socio-technical enabler of resilience, inclusion, and empowerment. Ultimately, the project illustrates that the future of XR lies not only in technological excellence but in its ability to enhance human dignity, collaboration, and trust.

25. Legal and Ethical Issues of SUN XR

Cong Yao [1] *and Paul Quinn* [1]

[1] Vrije Universiteit Brussels (VUB), Belgium

Abstract. This chapter provides a comprehensive analysis of the legal, ethical, and societal issues inherent in the SUN project's use of Extended Reality (XR) technologies. It examines the implications of processing personal and sensitive data collected via wearable devices, 3D acquisition, and human-machine interaction technologies under the European data protection framework, particularly the GDPR. The chapter further explores key ethical challenges, including obtaining informed consent from vulnerable participants, ensuring transparency in artificial intelligence systems, and balancing fundamental rights with the project's vital interests. By integrating ethics-by-design and privacy-by-design principles, the chapter outlines a framework for the responsible development and deployment of XR technologies, aiming to build societal trust and ensure compliance with relevant legal and ethical standards.

25.1 Introduction

This Chapter aims to describe the legal and ethical issues of the project, ensuring compliance with relevant regulations such as the European Data Protection Framework and the Medical Device Regulation. A broad range of legal and ethical requirements applicable to the project is addressed in this Chapter. Some of the key aspects considered, but not limited to, are:

- The data protection framework, including the General Data Protection Regulation (GDPR). As far as possible, key national provisions linked to data protection have also been considered;

- The ethical requirements incumbent upon the SUN platform in general, including the use of data in the pilots, and the possible contexts foreseen.

25.2 The Right to Privacy and Data Protection

As the SUN project involves the use of a series of mobile, wearable, and sensorial technologies, a large volume and various categories of personal data are processed. It is important to consider how these processing activities affect the fundamental rights of individuals involved, in particular the right to privacy and the right to data protection.

25.2.1 The right to privacy

The authors Samuel Warren and Louis Brandeis established the right to respect one's private life as a distinct notion in 1890 [Warren and Brandeis 1890], pp.193-220. The concept arose in response to technological advancements of the time, specifically instantaneous photographs and newspaper enterprises, and their increased negative impact on people's privacy through the collection and publication of unauthorised portraits of private individuals or for commenting on private and domestic life affairs [Roda et al. 2020], pp.10.

Although the concept of privacy has been around for over a century, there is no single, widely agreed-upon definition. The term "privacy" is highly dependent on several social, ethical, cultural, and other perspectives, as well as circumstances. Daniel Solove classified these approaches and theories on privacy into six categories [Solove 2010], ch.2, pp.2:

- The right to be let alone, that is *"to live one's life as one chooses, free from assault, intrusion or invasion except as they can be justified by the clear needs of community living under a government of law"*;[1]
- The limited access to the self, as the ability to shield oneself from unwanted public observation and discussion by others;
- Secrecy, where privacy is infringed by public disclosure of previously concealed information and where the interest of the individual is to avoid disclosure of personal matters;[2]

[1] Justice Abe Fortas as cited in [Solove 2010], ch.2, pp.2
[2] Whalen v. Roe (1977) as cited in [Solove 2010], ch.2, pp.5

- Control over personal information, meaning the claim of individuals, groups, or institutions to determine how, when, and to what extent information about them is given to others;[3]
- Personhood, concerns the protection of the integrity of personality and is considered to be *"those attributes of an individual* which are irreducible in the selfhood" [Solove 2010], pp.9;
- Intimacy, where the focus is on the development of personal relationships and different degrees of intimacy and self-revelation [Roda et al. 2020], pp.10.

Respect for private life and 3D acquisition technologies

In the SUN project, 3D acquisition technologies can capture highly detailed and accurate representations of objects or individuals in three dimensions. However, it is important to acknowledge that such technologies may be considered intrusive by some individuals. Privacy concerns arise when individuals are not aware of the data being captured or how they are utilized. The data collected from the physical environment or human beings through 3D technologies could potentially be misused for unauthorized purposes, including identity theft or unwarranted surveillance. Therefore, it is crucial to establish and implement robust privacy safeguards when utilizing 3D technologies. These safeguards ensure that individuals' privacy rights are respected and that data is handled securely and responsibly. By doing so, the project mitigated potential risks and uphold ethical standards in the use of 3D acquisition technologies.

Respect for private life and human-machine interaction technologies

Human-Machine Interaction (HMI) technologies have the capability to collect vast amounts of data every second in order to evaluate users' performance and enhance safety. This data encompasses various types, including biometric data, health data, behavioural patterns, location data, and user preferences. However, it is important to recognize that these data points can be sensitive and reveal information about individuals' behaviours and characteristics. Furthermore, the type and volume of data collected are sufficient to create detailed profiles of XR device users. If used for unintended purposes, such data may intrude upon the private lives of various individuals, including XR users, healthcare service providers, and bystanders.

In the SUN Pilots, wearable devices are employed to monitor users' health conditions (e.g., in the XR for rehabilitation Pilot), daily routines (e.g., in the XR for work Pilot), well-being, and work performance. These devices aim to assist users in leading healthier

[3]Alan Westin as cited in [Solove 2010], ch.2, pp.5

and safer lifestyles by providing insights into how their bodies respond, move, and rest, enabling them to adopt systematic and personalized guidance. For instance, in the XR for rehabilitation Pilot, wearable XR devices collect health data such as blood pressure and heart rate to evaluate users' health status, allowing the XR system to offer specific movement guidance accordingly. During the use of XR devices, users generate substantial amounts of data, gradually becoming repositories of information. These data may contain sensitive details that reflect users' private lives, as well as the lives of others. Additionally, these devices can collect information not only about the user but also about bystanders, who could be strangers, children, intimate partners, or anyone else.

The collection and usage of such data necessitate careful consideration of privacy safeguards and ethical practices to ensure the protection of individuals' privacy and the responsible handling of data. It is essential to implement robust measures to secure and handle this data with the utmost confidentiality, consent, and respect for individual rights. Transparent communication and informed consent have been prioritized to ensure that users and other individuals are fully aware of the data being collected and how they are used. Striking the right balance between technological advancement and safeguarding privacy is crucial in the development and deployment of human-machine interaction technologies [Pahi and Schroeder 2023].

Respect for private life and real-time monitoring

While the collection and processing of data by wearable XR devices may be legal, concerns do exist regarding the potential monitoring of users and interference with their private lives. As users move in both private and public spaces, XR devices generate data that enables the observation and evaluation of user behaviours and performance, ultimately leading to personalized guidance. However, this observation and evaluation can be intrusive when XR devices collect data that reveals aspects of individuals' private lives, such as their preferences, religion, sexual orientation, and more.

Real-time monitoring technologies in XR devices may rely on geographical positioning transponders that operate with a high level of accuracy. A centralized communication system can be utilized to track the routes of each XR device. XR manufacturers and end-user organizations may potentially access this private information to enable various functions. While such data may be crucial for the performance and improvement of XR devices, such as in the XR for rehabilitation Pilot, they can also be highly sensitive as they have the potential to reveal the private lives of XR device users. Membership in religious or political organizations and sensitive personal relationships could be disclosed to various stakeholders.

The impact of real-time monitoring technology is not limited to end users alone; it also poses privacy risks to bystanders. For instance, facial images, verbal communications, and precise locations of bystanders can be collected when XR devices are in use. In this context, the privacy risks to bystanders from the XR technology's database are comparable to or even higher than those faced by end users. Bystanders often face challenges in exercising their rights over their personal data because they may be unaware that their information is being collected. Therefore, technological measures have been implemented to proactively anonymize bystander data through techniques such as blurring or providing selective options to avoid excessive collection of bystanders' information [Pahi and Schroeder 2023], pp.12.

By implementing privacy-enhancing technologies and practices, the SUN project minimized the privacy risks associated with real-time monitoring and ensured the protection of both end users and bystanders. Anonymization techniques and clear communication regarding data collection and usage can help mitigate these risks and preserve privacy rights in the deployment of XR technologies.

25.2.2 The right to Personal Data Protection

The protection of natural persons in relation to the processing of personal data is a fundamental right provided by the GDPR [GDPR 2016]. At the European level, legal protection of personal data is supported by Article 8 of the European Convention on Human Rights (ECHR) and in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data No 108 [ETS No. 108 1981]. At the European Union level, it is provided by Article 8(1) of the Charter and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU), stating that everyone has the right to the protection of personal data concerning him or her [ETS No. 108 1981].

The right to the protection of personal data is closely linked to the right to privacy. They are considered vital components of a sustainable democracy [ETS No. 108 1981]. The notion of data protection originated from the right to privacy, and they share similar values. Both rights are instrumental in protecting and promoting fundamental values and rights, and in exercising other rights and freedoms, e.g., the freedom of expression or the right to assembly. Besides, in the EU, neither are absolute rights, and they can be limited under certain conditions under the EU Charter of Fundamental Rights. They must be balanced against other EU values, human rights, or public and private interests, such as freedom of expression or access to information. They also need to be weighed up against other public interests, such as national security [EDPS 2023].

Simultaneously, the EU recognizes privacy and data protection as two separate rights. The right to privacy or the right to a private life plays a pivotal role, which addresses the control of information about oneself, being autonomous, and being let alone. Privacy is not only an individual right but also a social value. Data protection aims to ensure the fair processing of personal data by both the public and private sectors. It addressed the protection of any information relating to an identified or identifiable natural (living) person. Hence, data protection is more about an individual's right compared to privacy.

Processing of personal data with the use of 3D acquisition Technologies

3D acquisition technologies can raise data protection issues due to their involvement in collecting and processing personal data. These technologies can capture highly detailed and accurate images of physical objects, including people's faces and bodies, resulting in the creation of 3D models that contain personal data. The use of 3D acquisition technologies for biometric identification or surveillance purposes can potentially infringe upon individuals' privacy and data protection rights, thereby necessitating stricter data protection regulations to prevent the misuse of this sensitive data. Another data protection concern is the unauthorized collection, use, and sharing of personal data. For instance, if a 3D scanner is utilized in a public space, individuals in the vicinity may not be aware that their images are being captured and used to generate 3D models. This lack of awareness makes it more challenging for data subjects to exercise their right to personal data protection. Furthermore, there is a risk of data breaches or cyberattacks compromising the personal data collected through these technologies. Once captured, 3D data may be stored on servers or other digital devices, and without adequate security measures in place, it becomes susceptible to theft or misuse. Therefore, SUN's partners utilizing 3D acquisition technologies, implemented appropriate data protection measures, including obtaining informed consent from individuals prior to collecting their personal data, implementing robust security measures to safeguard the data, and ensuring compliance with relevant data protection laws and regulations.

Processing of personal data with the use of human-machine interaction technologies

Human-Machine Interaction (HMI) technologies collect various data about users for functional purposes, such as ensuring smooth user movements within the XR system and effective communication within the system. However, these data sets can also be utilized for multiple other purposes, which can pose risks to privacy and data protection.

One of the key data protection concerns with HMI technologies is the potential for unauthorized collection and use of personal data. Given the volume of data collected, users may have limited control or knowledge of all the data points being collected and processed. Data controllers would then have the freedom to use this inferred

data for various purposes, including targeted advertising or workplace discrimination. Additionally, if a virtual assistant is constantly listening to a user's conversations, there is a risk of inadvertently capturing sensitive information that the user did not intend to share.

Transparency around data collection and processing by HMI technologies is another significant issue. Users may lack awareness of what data is being collected, how it is being used, and who has access to it. This lack of transparency hinders users' ability to make informed decisions about their privacy and data protection. Therefore, SUN's partners involved in developing and deploying HMI technologies implemented appropriate data protection measures, including clear and transparent information about data collection and processing, obtaining informed consent from users, and ensuring compliance with relevant data protection laws and regulations.

Processing of personal data with the use of artificial intelligence capacities

The data collected in the SUN project is analyzed using Deep Learning algorithms, enabling the system to comprehend the user's health and emotional status progression. The objective of the SUN system is to generate personalized models for each user, enabling the detection of abnormalities by identifying deviations from expected behaviour. By employing AI and health analytics techniques, the system leverages the analysed data to provide personalized recommendations for the user's rehabilitation plan in the XR for Rehabilitation Pilot.

AI algorithms may underlie automated decision-making and profiling [Article 29 Working Party 2018] and invoke the GDPR application. In the context of the GDPR, profiling, a form of automated decision-making, is defined as the use of personal data to evaluate personal aspects relating to a natural person, such as their performance, health, or behaviour [Article 29 Working Party 2018].

The Guidelines on Automated Decision-Making and Profiling note that "Automated decision-making has a different scope and may partially overlap with or result from profiling" and that "Automated decisions can be made with or without profiling; profiling can take place without making automated decisions". They further specify that "Solely automated decision-making is the ability to make decisions by technological means without human involvement"[Article 29 Working Party 2018].

Controllers are generally allowed to engage in profiling and automated decision-making as long as they adhere to the relevant data processing principles and have a lawful basis for the processing. However, additional restrictions and safeguards apply specifically to solely automated individual decision-making. According to Article 22 of the GDPR, unless an exception specified in subparagraph (2) or (4) applies, data

subjects have the right not to be subject to a decision based solely on automated processing, including profiling, if it produces legal effects or significantly affects them. One exception is when the data subject has provided informed consent [GDPR 2016], Art.22.

In relation to sensitive data, automated individual decision-making may only be allowed if the legal basis for processing is explicit consent or substantial public interest. As stated before, SUN's partners involved in developing and deploying artificial intelligence capacities implemented appropriate data protection measures, including clear and transparent information about data collection and analysis, obtaining informed consent from users, and ensuring compliance with relevant data protection laws and regulations.

25.3 Ethical and Societal Concerns

The SUN project raises several ethical and societal concerns, primarily stemming from the power imbalance between data controllers and XR users, necessitating robust safeguards to protect fundamental rights and prevent exploitation, especially for vulnerable groups. Promoting transparency, accountability, and responsible innovation is essential to building societal trust and ensuring that the technologies align with public values, which can be achieved through active stakeholder engagement. Furthermore, the project must continuously balance fundamental rights like privacy and data protection with its vital interests and potential benefits, implementing adequate measures to safeguard individuals while enabling positive research outcomes.

25.3.1 Sources for Principles of Ethics in Research with Humans

The SUN project involved research engaging directly with human participation in each Pilot. Some Pilots, for example, XR for rehabilitation and XR for communication, involve research for medical devices. Hence, several principles should be followed during the research with human participation, considering the interests of individuals involved in the project.

Medical ethics is one of the most important principles for any research dealing directly with human participants. These ethical principles have been codified in various instruments nationally and internationally. For example, one such instrument is the Declaration of Helsinki, first adopted by the World Medical Association in 1964 and subsequently amended, which was adopted as a statement of ethical principle for medical research involving human subjects, including research on identifiable human

material and data [WMA 2013], Preamble, para. 1. While the Declaration of Helsinki is mainly aimed at physicians, it encourages others involved in medical research with human participation to adopt these principles [WMA 2013], para. 2. It includes guiding principles related to risks, burdens, and benefits for human participants in research, vulnerable groups and individuals, informed consent, confidentiality, and research ethics committees.

25.3.2 Informed Consent

Informed consent is a cornerstone of the principle of autonomy and is relevant to research with human participants. The Declaration of Helsinki provides that "after ensuring that the potential subject has understood the information, the physician or another appropriately qualified individual must seek the potential subject's freely-given informed consent, preferably in writing" [WMA 2013], para. 26. The Declaration of Helsinki also lists the sort of information that needs to be provided to the research participants for the consent to be informed. It requires that special attention is given "to the specific information needs of individual potential subjects as well as to the method used to deliver the information" [WMA 2013].

It is also a central element of both the Oviedo Convention [ETS No. 164 1997], Art. 5 and the Oviedo Additional Protocol [ETS No. 165 2005], Art. 15 Article of the Oviedo Convention sets out the conditions for undertaking research on a person, including that "the necessary consent as provided for under Article 5 has been given expressly, specifically and is documented. Such consent may be freely withdrawn at any time" [ETS No. 164 1997], Arts 5-6. The importance of informed consent in research with human participants is further evidenced by its inclusion in various instruments, including the ICCPR [UN 1966], CIOMS Guidelines [CIOMS 2016], pp. 33, the ICH GCP [ICH 1996], Principles 2.9-2.10, pp. 15-18, the WHO GCP [WHO 1995], Principle 7, pp. 59-71, and the UNESCO Declaration [UNESCO 2005], art. 6.

When considering these requirements related to informed consent in the context of the SUN project, it is important to consider that each Pilot in the SUN project involves research with human participants. Those participants include patients, employees, and people with communication disabilities. Not all participants would likely be able to give consent for their participation in the Pilots and the processing of their personal data themselves. Accordingly, the participants in the SUN project could be separated into two groups: those who are able to sign informed consent related to their participation and the processing of their personal data, and those who might not be able to give their consent because of a lack of competent judgment. Traditionally, the following elements are considered necessary for competent judgment: the ability to receive,

process, and understand information, the ability to appreciate the situation and its consequences, the ability to weigh benefits, risks, and alternatives, and the ability to make and communicate a decision.

Furthermore, considering those people who cannot give their consent for their participation in the Pilots and the processing of their personal data, additional consideration shall be given to the inclusion of vulnerable groups in the SUN Pilots. Vulnerable persons are described as those who are relatively or absolutely incapable of protecting their own interests [UNESCO 2005], Principle 7, pp. 65. While it is recommended not to label a member of a certain group as vulnerable automatically, some characteristics make it reasonable to assume that certain individuals are vulnerable [CIOMS 2016], pp. 15, for instance, persons in nursing homes, those incapable of giving consent, or with diminished mental capacities, people with incurable disease, people with physical frailty (e.g., due to age or co-morbidities), children, or economically disadvantaged persons. It is recommended to make the determination of whether a participant is to be considered a vulnerable person based on the specific context of their case.

25.3.3 Ethical Artificial Intelligence

One of the key technological challenges in the SUN project is the integration of artificial intelligence (AI) for data analysis obtained from XR device users. AI techniques are applied to process raw scans, sparse 3D points, or incomplete acquisitions, enabling the generation of high-quality 3D triangular meshes. Additionally, AI is used to explore unknown environments within resource-constrained scenarios. Neural generative models enhance image resolution while ensuring seam continuity, and they also contribute to generating complete scene 3D renderings by extrapolating context-based information that aligns with semantic and geometric constraints.

The main concern of AI's application is its "black box" character, which leads to a lack of transparency and accessibility, leading to a decline in trust [Kiseleva 2019]. AI algorithms are based on deep machine learning, a fast, automatic, data-hungry self-learning mechanism [Coglianese and Lehr 2016]. However, deep machine learning is not inherently explainable, making it challenging to predict and explain the outputs of AI systems. This lack of predictability and understandability reduces human control and comprehension throughout the decision-making process, resulting in autonomy and the black-box effect as prominent characteristics of AI, which pose challenges for its application [Kiseleva 2019].

Nevertheless, legislation, such as the GDPR, mandates transparency in the processing of personal data based on automated decision-making. The GDPR requires that individuals are provided with meaningful information regarding the logic behind auto-

mated decision-making, as well as the significance and consequences of such decisions. These rights aim to address the transparency issue associated with AI applications, mitigating concerns arising from the autonomy and black-box characteristics of AI systems [Kiseleva 2019].

Concerning the ethical issue of AI, a key concept proposed by the European Commission is trustworthy AI. The main ethical principles to achieve Trustworthy AI are provided in the Ethics Guidelines on Trustworthy AI issued by the High-Level Expert Group on AI (AI HLEG) on 8 April 2019.[4] The Expert Group considers that AI has the potential to transform society significantly and would be a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation [AI HLEG 2019], pp.4. Under the Guidelines, three key components of trustworthy AI are addressed:

- *Lawful*, complying with all applicable laws and regulations;
- *Ethical*, ensuring adherence to ethical principles and values; and
- *Robust*, both from a technical and social perspective, since even with good intentions, AI systems can cause unintentional harms [AI HLEG 2019], pp.5.

Besides, the Guidelines also state that trustworthy AI should respect the following four principles [AI HLEG 2019], pp.11-12:

- *Respect of human autonomy*: humans interacting with AI systems must be able to keep full and effective self-determination over themselves and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition, or herd humans. Instead, they should be designed to augment, complement, and empower human cognitive, social, and cultural skills;
- *Prevention of harm*: AI systems should neither cause nor exacerbate harm[5] or otherwise adversely affect human beings.[6] AI systems and environments must be safe and secure. They must be technically robust, and it should be ensured that they are not open to malicious use. Vulnerable persons should receive greater attention during the application of AI systems;

[4]AI HLEG is an independent expert group that was set up by the European Commission in June 2018.
[5]Harms can be individual or collective, and can include intangible harm to social, cultural, and political environments.
[6]This entails the protection of human dignity as well as mental and physical integrity. This also encompasses the way of living of individuals and social groups, avoiding, for instance, cultural harm.

- *Fairness*: the development, deployment, and use of AI systems must be fair;[7]
- *Explicability*: processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions - to the extent possible - explainable to those directly and indirectly affected.

While some of these ethical principles are also reflected in legal requirements, thereby falling into the scope of lawful AI, it is significant to recall that adherence to ethical principles "goes beyond formal compliance with existing laws." [AI HLEG 2019], pp.12.

To operationalize these principles, SUN project partners established clear protocols for human oversight, algorithmic fairness, and accountability, ensuring the AI systems used in the project are both effective and ethically sound.

25.3.4 The Necessity to Balance between Fundamental Rights and Vital Interests of Different Groups of People

During the Pilots of the SUN project, XR devices collect data in various settings, including homes, hospitals, rehabilitation rooms, and public places. These XR environments rely on the interplay of multiple sensors, large volumes and varieties of data, and various algorithms and AI systems. While the data collected is essential for the functioning of the technologies and the interests of the project partners, it also poses risks to the fundamental rights of both users and bystanders. Therefore, it is crucial to strike a balance between fundamental rights and the vital interests of different individuals.

When users use XR devices in the XR system, there is an inherent risk that sensors or mobile devices can record the activities of not only the users but also other people who are not directly involved in the Pilots. These individuals may include family members of users, staff, or visitors in hospitals or rehabilitation rooms, and even bystanders in public places. Although the mere collection of activities from these individuals does not necessarily mean processing their personal data, it can still affect their fundamental rights in certain cases. Moreover, in workplace pilots, obtaining genuine 'free consent' from employees can be challenging due to inherent power imbalances. Staff may feel unable to refuse or withdraw consent without fear of repercussions, necessitating additional protective measures beyond simple consent forms.

[7]Fairness has both a substantive and a procedural dimension. The substantive dimension implies a commitment to ensuring equal and just distribution of both benefits and costs and ensuring that individuals and groups are free from unfair bias, discrimination, and stigmatisation. The procedural dimension of fairness entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them.

25.4 Conclusions: : Building a Responsible Future for XR

The SUN project stands at the forefront of technological innovation with immense potential to improve human health, safety, and well-being. However, this potential cannot be realized without a steadfast commitment to navigating the accompanying legal, ethical, and societal complexities.

This chapter has described what SUN did and has outlined a framework for future actions. Success depends on moving beyond mere regulatory compliance and embracing a holistic strategy of ethics-by-design and privacy-by-design. This means:

- Embedding safeguards into the technology from the earliest stages of development;
- Fostering transparent communication with users, participants, and the public about how data is used;
- Establishing clear accountability structures within the consortium;
- Continuously monitoring and evaluating the societal impact of the Pilots.

The choices made by the SUN project not only determine its own success but also contribute to setting crucial precedents for the responsible development of XR technologies globally. By prioritizing human rights and societal trust, SUN can ensure that its legacy is one of both groundbreaking innovation and unwavering ethical integrity.

References

Article 29 Working Party (2018). *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679. 17/EN WP251rev.01*. Adopted on 3 October 2017; last revised and adopted on 6 February 2018. URL: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

Coglianese, Cary and David Lehr (2016). "Regulating by robot: Administrative decision making in the machine-learning era". In: *Geo. LJ* 105, p. 1147.

Council for International Organizations of Medical Sciences (CIOMS) (2016). *International Ethical Guidelines for Health-related Research Involving Humans*. Guideline 15. Council for International Organizations of Medical Sciences. URL: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf.

Council of Europe (ETS No. 108) (1981). *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*. European Treaty Series No. 108. Council of Europe. URL: http://www.conventions.coe.int/Treaty/en/Treaties/Html/108.htm.

Council of Europe (ETS No. 164) (1997). *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*. European Treaty Series No. 164. Oviedo: Council of Europe.

Council of Europe (ETS No. 165) (2005). *Additional Protocol to the Convention on Human Rights and Biomedicine Concerning Biomedical Research*. European Treaty Series No. 165. Oviedo: Council of Europe.

European Data Protection Supervisor (EDPS) (2023). *Data Protection*. Last accessed 15 Feb 2023. URL: https://edps.europa.eu/data-protection/data-protection_en.

European Parliament and the Council of the European Union (GDPR) (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, Recital 1*.

Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), International Council for (1996). "Guideline for good clinical practice E6 (R2)". In: *ICH Harmon Tripart Guidel* 1996.4.

High-Level Expert Group on Artificial Intelligence (AI HLEG) (Apr. 2019). *Ethics Guidelines for Trustworthy AI*. Published 8 April 2019. European Commission. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Kiseleva, A. (July 2019). "Decisions made by AI versus transparency: Who wins in Healthcare?" In: *The futures of eHealth. Social, Ethical and legal challenges*. Ed. by T. C. Bächle and A. Wernick. Berlin, Germany: Humboldt Institute for Internet and Society.

Pahi, Suchismita and Calli Schroeder (2023). "Extended privacy for extended reality: XR technology has 99 problems and privacy is several of them". In: *Notre Dame J. on Emerging Tech.* 4, p. 1.

Roda, Sara, Istvan Mate Borocz, Ioulia Konstantinou, Eike Gräf, Vicky Vouloutsi, Spyros Karamoutsos, Dirk Wollherr, Maider Arieta-araunabeña, Belén Garnica, Sixto Arnaiz, et al. (2020). "HR-Recycler deliverable D2. 1 Report on security, data protection, privacy, ethics and societal acceptance". In.

Solove, Daniel J (2010). *Understanding privacy*. Harvard University Press.

United Nations (UN) (1966). *International Covenant on Civil and Political Rights*. United Nations Treaty Series (UNTS). Adopted 16 December 1966, entered into force 23 March 1976, 999 *UNTS* 171, art. 7.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2005). *Universal Declaration on Bioethics and Human Rights*.

Warren, Samuel and Louis Brandeis (1890). "The right to privacy". In: *Killing the messenger: 100 Years of media criticism*. Columbia University Press, pp. 1–21.

World Health Organization (WHO) (1995). *Guidelines for Good Clinical Practice (GCP) for Trials on Pharmaceutical Products*. Geneva.

World Medical Association (WMA) (2013). *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants (June 1964, most recently amended October 2013)*. Adopted 1964, amended 2013.

Extended Reality (XR) is a rapidly growing technology that bridges physical and virtual worlds, opening up new possibilities in healthcare, communications, and security. The European project SUN – Social and hUman ceNtered XR, funded by the Horizon Europe program, addresses the ongoing challenges of making XR more accessible, usable, and realistic. SUN develops technologies and models that enhance social interaction and immersive perception, while keeping an ethical and human-centered design, by introducing new wearable sensors, haptic interfaces, and high-performance streaming solutions. Through new 3D acquisition techniques and the use of artificial intelligence, SUN explores innovative ways to connect physical objects and digital counterparts, creating coherent and immersive environments. The project's innovations were validated in three real-world piloting scenarios: rehabilitation therapy, workplace safety and social interaction, and assistive technologies for individuals with severe mobility or communication impairments. This volume presents the results of three years of research and development, offering a solid vision of how XR can evolve in a sustainable, ethical, and human-centered way.