Social and hUman ceNtered XR

Claudio Vairo[1,*], Marco Callieri[1], Fabio Carrara[1], Paolo Cignoni[1], Marco Di Benedetto[1], Claudio Gennaro[1], Daniela Giorgi[1], Gianpaolo Palma[1], Lucia Vadicamo[1] and Giuseppe Amato[1]

[1]ISTI-CNR, via G. Moruzzi, 1, Pisa, 56100, Italy

Abstract

The Social and hUman ceNtered XR (SUN) project is focused on developing eXtended Reality (XR) solutions that integrate the physical and virtual world in a way that is convincing from a human and social perspective. In this paper, we outline the limitations that the SUN project aims to overcome, including the lack of scalable and cost-effective solutions for developing XR applications, limited solutions for mixing the virtual and physical environment, and barriers related to resource limitations of end-user devices. We also propose solutions to these limitations, including using artificial intelligence, computer vision, and sensor analysis to incrementally learn the visual and physical properties of real objects and generate convincing digital twins in the virtual environment. Additionally, the SUN project aims to provide wearable sensors and haptic interfaces to enhance natural interaction with the virtual environment and advanced solutions for user interaction. Finally, we describe three real-life scenarios in which we aim to demonstrate the proposed solutions.

Keywords

extended reality, artificial intelligence, digital twins

1. Introduction

The Social and hUman ceNtered XR (SUN) project aims at investigating and developing extended reality (XR) solutions that integrate the physical and the virtual world in a convincing way, from a human and social perspective. The virtual world will be a means to augment the physical world with new opportunities for social and human interaction. Figure 1 summarizes the current limitations and SUN's solutions to achieve this.

More in detail, the relevant limitations that we will address are:

- *Lack of solutions to develop scalable and cost-effective new XR applications*: building an XR application for a new physical environment requires creating from scratch its accurate digital twin, with both visual and physical/functional properties, with significant efforts and high costs.
- *Lack of convincing solutions for mixing the virtual and physical environment*: generally, augmented reality is limited to the visual alignment of physical and virtual components. However, a convincing XR requires physical and virtual elements to interact, and actions in one world to have an effect on the other world.
- *Lack of plausible and convincing human interaction interfaces in XR*: human interaction in XR should include natural interaction, where actions in the physical world have convincing counterparts in the virtual one. Similarly, objects manipulated in the virtual environment should provide users with a realistic feeling.
- *Barriers due to resource limitations of end-user devices*: XR end-user devices at an affordable cost often have limited computing, memory, and networking resources. Therefore, they cannot handle highly realistic and convincing 3D models that embed visual, semantic, and physical properties.

The SUN project will develop solutions to surpass current limitations, including:

- Scalable solutions to obtain plausible and convincing virtual copies of physical objects and environments
  - The system will learn from the physical world, during its use. Like a baby, who learns how to recognize and use objects with the experience, SUN will use artificial intelligence, computer vision, and sensor analysis to incrementally learn, during its usage, the visual and physical properties of real objects, and to generate, recognize, and use digital twins in the virtual environment.
  - Learned objects and environments will be incrementally added to the SUN Digital Twins Library of available items and reused in various XR applications.

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
claudio.vairo@isti.cnr.it (C. Vairo)
0000-0000-0003-2740-4331 (C. Vairo); 0000-0002-5832-680X (M. Callieri); 0000-0001-5014-5089 (F. Carrara); 0000-0002-2686-8567 (P. Cignoni); 0000-0001-5781-7060 (M. Di Benedetto); 0000-0002-3715-149X (C. Gennaro); 0000-0002-6752-6918 (D. Giorgi); 0000-0002-1032-2092 (G. Palma); 0000-0001-7182-7038 (L. Vadicamo); 0000-0003-0171-4315 (G. Amato)

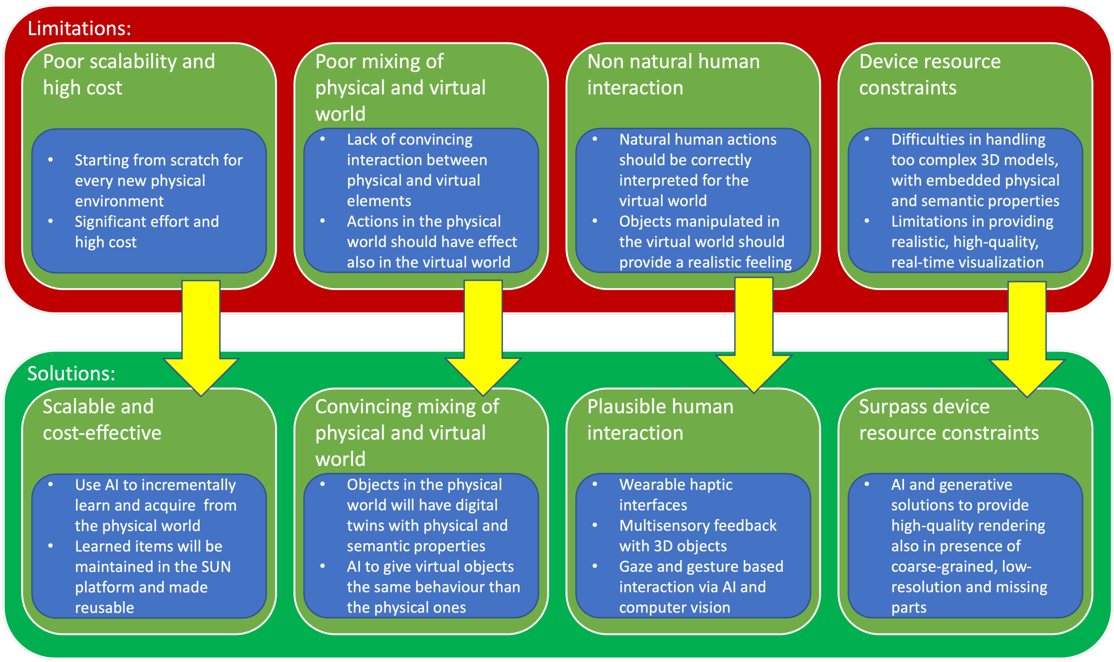

Figure 1: Limitations and solutions proposed by SUN.

- Seamless and convincing interaction between the physical and the virtual world
  - Objects and environments of the physical world have a digital twin in the virtual world, with the same physical and visual properties.
  - Manipulating an object in the physical world will have the same plausible effect in the virtual world. For instance, a physical wrench could be used to unscrew a virtual bolt. Complementarily, a virtual wrench manipulated in the virtual world, will provide the user with a feeling consistent with a physical wrench.
- Wearable sensors and haptic interfaces for convincing and natural interaction with the virtual environment
  - Wearable haptics will enhance physical interaction with virtual menus and contextual information displayed to the user. In addition, solutions for body contextual information (such as skin stretch-vibration on body parts) will lightly guide the user in tasks such as remote training, home physical exercises, and interaction with other persons.
  - Novel wearable haptic interfaces will allow multisensory feedback (vibrotactile and thermal cues under fingertips) for XR scenarios, such as interaction with 3D virtual objects.
  - We will provide advanced solutions for user interaction, including gaze-based and gesture-based interaction.
- Artificial intelligence-based solutions to address current computing, memory, and network limitations of wearable devices
  - AI will be used to generate plausible, high-quality renderings also for coarse-grained, low-resolution, or incomplete 3D models, leveraging on solutions similar to those used for deep-fake generation.

The solutions will be demonstrated in three real-life scenarios. The rest of the paper is organized as follows: Section 2 describes the scenarios. Section 3 presents the technical parts of the project. Finally, Section 4 concludes the paper.

2. Scenarios

The project will assess the developed technologies in these three real-life scenarios:

1. Extended reality for rehabilitation.

2. Extended reality for safety and social interaction at work.
3. Extended reality for people with serious mobility and verbal communication diseases.

XR experiments are going to be good examples to help better understand the social impact of those technologies even considering that COVID-19 boosted a positive attitude toward them [1]. The SUN application for safety at work, fragile people in rehabilitation, and in social cues communication can act as the driving forces behind developing the empathic stimuli and facilitating XR empathic experiences [2].

2.1. Extended reality for rehabilitation

The aim of the proposed rehabilitation scenario is to motivate the patient to exercise efficiently by providing feedback in relation to performance, while the physiotherapy exercises are performed in any setting, e.g. clinical, at home, indoors, outdoors, or even in public areas. This scenario is based on the use of a digital tool employing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) to assist and monitor individual motor learning in the context of a supervised personalized remote exercise rehabilitation program for the management of injuries/pathologies. The digital tool will also enable supervised personal training. Important success factors for home-based programs is to include patients with a favorable prognosis and increasing adherence to the program. The main goal of this scenario is to improve compliance to the physiotherapy protocol, increase patient engagement, monitor physiological conditions, and provide immediate feedback to the patient by classifying an exercise in real-time as correctly or incorrectly executed, according to physiotherapists' set criteria. It is already known that visuo-physical interaction results in better task performance than visual interaction alone tracking performance. Wearable haptics (such as EMG) can be used to enhance physical interaction by monitoring body contextual information. Visual AI algorithms can enhance the understanding of the alignment of the body, capturing the outline and giving real-time personalized feedback to improve the quality of the movement. Rehabilitation adherence and fidelity are especially challenging, alongside motor learning with personalized feedback. Physiotherapy after orthopedic surgery and other kinds of upper limb motor impairment such as "frozen shoulder" or severe hand arthritis is crucial for complete rehabilitation but is often repetitive, tedious, and time-consuming. Actually, in order to achieve motor recovery a very long physiotherapy treatment, sometimes more than 50 sessions, is needed. Therefore, there is a necessity to suitably address the evidence-practice gap and translate digital innovations into practice while enabling their improvement, enhancing accuracy as well as portability in the home environment. Starting from a visual representation of an exercise, delivered through an avatar of the therapist performing it in the correct way, the patient will be called to repeat it, or in sync with the avatar performing the exercise. Wearables and wireless sensors will be used to measure and deliver contextual information, including multiple inertial measurement units (dynamic motion, body and limb orientation kinetics & kinematics, maximum respective joint angle achieved), surface electromyography (neuromuscular potentials such as activation and fatigue), and smartwatch (heart rate, oxygen saturation, arterial blood pressure). Besides an accelerometer, a flex sensor could be employed instead of a gyroscope for specific angle measurement during limb bending and a force-sensitive sensor to measure limb pressure, with the potential of classifying a variety of rehabilitation exercises over a sensor network. Data from a camera and the sensors will be collected in real-time and send to a GPU server (either cloud-based exploiting the potential of cloud-based AI infrastructure or edge-based exploiting edge AI) [3, 4]. An AI algorithm will process the sensory input, responding to the level of muscle activation by reducing or enhancing the exercise difficulty-intensity, and to the movement quality providing personalized feedback for movement correctness. Clinicians' feedback can also be added at any point during the automated feedback to enhance accuracy, secure safety and further improve the algorithm. Usability testing is required a priori to the application of such complex innovative digital tools. The goal of this scenario is to also increase user engagement for diverse populations. These can be achieved by the notion of personalization for diverse populations and rehabilitation needs and abilities. An iterative-convergent-mixed-methods design will be used to assess and mitigate all serious usability issues and to optimize user experience and adoption. This design will provide transparency and guidance for the development of the tool and its implementation into clinical pathways. The methodological framework is defined by the ISO 9241-210-2019 (Human-centred design for interactive systems) constructs, effectiveness, efficiency, satisfaction, and accessibility.

2.2. Extended reality for safety and social interaction at work

AR and VR can create more immersive experiences for people at work in order to make their job safer, by providing new ways to be aware of possible hazards and receive more effective, engaging and entertaining training on safety procedures. This is to alert and prevent serious accidents provoked by the co-occurrences of different causes, which can be avoided by conscious collaboration. With VR/AR headsets, workers will be able, for example, to better understand difficult-to-grasp concepts or topics such as protocols and procedures for safety and security. It offers a great opportunity for the industry to be able to make use of a variety of different options relying on extended reality experiences to provide each worker with an immersive access to relevant information, and to encourage the adoption of safer behaviours. Extended reality encompassing virtual, augmented, and mixed reality brings immersive experiences to workers no matter where they are. A XR experience can "come alive" when a worker puts on a VR headset and walks, for instance, on a shop floor. Workers can experience hard-to-conceptualise current-day topics through extended reality, such as moving around in potentially dangerous environments. XR technology will improve effectiveness and user engagement. Through an adequate design of XR contents, immersive experience seamlessly integrates several principles not present when using non-immersive contents such as better contextualization and real-time decision-making feedback, in a safe yet realistic environment for practice. The Holo-Light's software Engineering Space AR 3S, which allows workers to visualise and work with design files including CAD in an XR environment, will be optimised and upgraded to encourage social interaction among workers to also provide them with alerts and indications about the possibility of incidents. Correct and incorrect behaviour will be simulated. Overall, it will enable designing, prototyping, quality assurance, and overall technical industrial training and education in emerging manufacturing technology scenarios also with the opportunity to not just increase the speed in acquiring new tasks or to improve industrial processes but overall, also increasing the safety for workers dealing with complex machines. Features that will be exploited in the scenario include:

- High Quality 3D Content in XR: To securely import and view all your design files for AM in XR. Visualize and manipulate even data-intensive 3D content.
- Merge the Real and Virtual: Place your CAD/engineering designs in the real world. Manipulate projected assemblies directly at their envisaged destination.
- Work on Complex Holograms: Fully manipulate your designs and assemblies. Place, rotate, adjust, resize, slice and dice. Save your work.
- Share the Experience: Collaborate with your colleagues, partners, or customers in AR. Set up local and global meetings in an XR environment.

2.3. Extended reality for people with serious mobility and verbal communication diseases

Some people with various motor disabilities or after strokes have huge difficulties in communicating with each other and even to address their vital needs. The project will join the challenge to find a dedicated communication pathway for those people introducing the possibility to interact with some specific social cues and transform them in clear communication or actions. The proposal is to count on residual abilities giving them a meaning in terms of communication supported when needed by avatars in a virtual environment. The objective is to realise low-cost non-invasive tools based for instance on the existing biofeedback, face expression and other input. The challenge will be to create a solution allowing the design of person specific extended reality interaction even at home in the working place or at school and developing novel multi-user virtual communication and collaboration solutions that provide coherent multisensory experiences and optimally convey relevant social cues. Successful implementation of the pilot for this scenario allows people with communication and motor disabilities to interact with friends and relatives, meeting realistically in an extended (physical + virtual) environment. The person with communication and motor disabilities will be represented by an avatar which will interact with other people staying in a physical augmented environment. Simultaneously, the person with communication and motor disabilities will have the illusion to be also in the same physical environment with the other people, and can interact with the new interfaces offered by SUN. SUN, in fact, will develop a new generation of non-invasive bidirectional body-machine interfaces (BBMIs), which will allow a smooth and very effective interaction of people with different types of sensory-motor disabilities with virtual reality and avatars. The BBMI will be able to decode motor information by using arrays of printed electrodes to record muscular activities and inertial sensors. The position of the sensors will be customised according to the specific motor abilities of the subjects. For example, it could be possible to use shoulder and elbow movements, and it could be also possible to record muscular activities from auricular muscles. All these sensors will be used to record information during different upper limb and hand functions movements. We will use a procedure to identify the signals more useful for the different tasks and implement a dedicated decoding algorithm. This approach allowed developing a simplified and yet very effective approach to control flying drones and will probably provide very interesting results also in this case. A machine learning approach will be implemented for the decoding of the different tasks relying on human-machine interfaces and in general on decoding information from electrophysiological and biomechanical signals. SUN will also allow to give sensory feedback to the user about the movement and the interaction with the avatar in the virtual environment, using transcutaneous electrical stimulation or small actuators for vibration.

3. Technical Challenges

In this section, we will describe in more detail the technical parts of the projects and the challenges that our institute of the CNR (ISTI) wants to address.

3.1. AI to learn objects for the virtual world

Various techniques exist that allow recognizing physical objects and environments and linking/registering them with the corresponding virtual ones to enable task-specific interactions between users and augmented environments. Existing techniques are based on image classification, object detection, and semantic segmentation via deep neural networks [5]. However, identifying, mapping, linking, and registering every single (instance of) physical object in the virtual context is often not a completely automatic process and requires manual effort.

In the context of the project, we will advance the current state of the art by developing AI-based solutions that allow the incremental and automatic discovery of environments, objects, and object usage patterns, while the system is being used. This will permit the incremental creation of libraries of reusable visually recognizable virtual objects and environments, the cost reduction in developing extended reality applications, and the scalability of extended reality to large scenarios.

3.2. Acquisition of physical properties for 3D objects

Currently, most of the efforts to create digital twins representation focus on the recreation of the shape and appearance of an existing object. However, the mechanical characteristics of a digital representation drive their actual physical behavior and are necessary for interactive environments to maintain the perceived plausibility of the representation. Obtaining such information practically and intuitively is a technical challenge [6]. Current solutions require laboratory settings or cumbersome devices [7]. We aim to face this challenge to create a digital replica of objects that support truly realistic interaction in XR environments, recreating convincing experiences.

The main objective is to provide innovative techniques to acquire and estimate the mechanical properties (like mass distribution or stiffness/softness) of 3D objects. This will be done by integrating wearable fingertip-sensing devices with maker-based video tracking during controlled object manipulation. AI data interpretation, followed by multimodal data fusion, will allow learning of the expected inertia tensor and other physical properties that best match the expected sensor readings. An important aspect is the design and 3D printing of objects with controlled mechanical properties for the acquisition of datasets with ground truth data to use during the learning phase.

3.3. AI in support of 3D completion and convincing XR presentation

XR environments are populated by 3D models representing real-world objects (assets). Assets coming from stock sources (commercial or open repositories) are often used to cope with the general situation, but for specific scenarios, it is often necessary to produce on the fly new 3D models. This task will address this phase of asset creation.

A first direction to pursue is to help in producing more complete and convincing 3D models by adding AI processing to the different steps of the acquisition pipeline. To this aim, the photogrammetric reconstruction seems the most promising one: AI processing can be effectively used in removing from photos those problems that reduce the accuracy of alignment and 3D data generation or unwanted areas [8]. Furthermore, AI can be used to fill gaps in the geometry and texture, like in [9], possibly having as data source the input photographic campaign, to ensure a more coherent filling.

On a different level, for specific classes of objects, we can exploit the a priori knowledge of the domain, to pre-train networks on examples. This will ensure better support in the 3D model generation and a more coherent and realistic completion of the models' geometries and textures.

3.4. AI-assisted 3D acquisition of unknown environments with semantic priors

Automatic environment discovery is a challenging task that has been revamped in the deep learning era, and automatic model instantiation has similarly received academic and industrial attention [10]. A unified solution for parallel exploration and semantic-driven user presentation is not yet available.

We will use reinforcement learning for automatic exploration (e.g., by unmanned vehicles) of unknown and possibly large objects and environments, while minimizing the scenario discovery time. To keep acquisition feasible in a time-constrained context and allow an immersive experience from the very start of the scenario exploration, our solution will also deploy pre-built 3D model instances for objects whose complete reconstruction would need to acquire information from different points of view, or explore novel methods (e.g., conditional denoising diffusion probabilistic models) to autogenerate the underlying representation.

4. Conclusions and Future Work

The Social and Human-centered XR project aims to overcome the current limitations in developing XR solutions that integrate the physical and virtual world from a human and social perspective. The limitations addressed in the SUN project include lack of solutions to develop scalable and cost-effective new XR applications, lack of convincing solutions for mixing the virtual and physical environment, lack of plausible and convincing human interaction interfaces in XR, and barriers due to resource limitations of end-user devices. The SUN project proposes scalable solutions to obtain plausible and convincing virtual copies of physical objects and environments, seamless and convincing interaction between the physical and virtual world, wearable sensors and haptic interfaces for convincing and natural interaction with the virtual environment, and artificial intelligence-based solutions to address current computing, memory, and network limitations of wearable devices.

Acknowledgments

This work has received financial support from the Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612 (Social and hUman ceNtered XR - SUN project) and by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Union under the NextGeneration EU programme.

References

[1] C. Kerdvibulvech, L. Chen, The power of augmented reality and artificial intelligence during the covid-19 outbreak, in: HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22, Springer, 2020, pp. 467–476.

[2] V. Paananen, M. S. Kiarostami, L.-H. Lee, T. Braud, S. Hosio, From digital media to empathic reality: A systematic review of empathy research in extended reality environments, arXiv preprint arXiv:2203.01375 (2022).

[3] L. Ciampi, C. Gennaro, F. Carrara, F. Falchi, C. Vairo, G. Amato, Multi-camera vehicle counting using edge-ai, Expert Systems with Applications 207 (2022) 117929.

[4] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Vairo, Facial-based intrusion detection system with deep learning in embedded devices, in: Proceedings of the 2018 International Conference on Sensors, Signal and Image Processing, 2018, pp. 64–68.

[5] L. Nicholson, M. Milford, N. Sünderhauf, Quadricslam: Dual quadrics from object detections as landmarks in object-oriented slam, IEEE Robotics and Automation Letters 4 (2018) 1–8.

[6] N. F. Duarte, K. Chatzilygeroudis, J. Santos-Victor, A. Billard, From human action understanding to robot action execution: how the physical properties of handled objects modulate non-verbal cues, in: 2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), IEEE, 2020, pp. 1–6.

[7] L. Marechal, P. Balland, L. Lindenroth, F. Petrou, C. Kontovounisios, F. Bello, Toward a common framework and database of materials for soft robotics, Soft robotics 8 (2021) 284–297.

[8] A. Murtiyoso, P. Grussenmeyer, Automatic point cloud noise masking in close range photogrammetry for buildings using ai-based semantic labelling, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-2/W1-2022 (2022) 389–393. doi:10.5194/isprs-archives-XLVI-2-W1-2022-389-2022.

[9] A. Maggiordomo, P. Cignoni, M. Tarini, Texture inpainting for photogrammetric models, Computer Graphics Forum in press (2023). URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14735. doi:10.1111/cgf.14735.

[10] A. Avetisyan, T. Khanova, C. Choy, D. Dash, A. Dai, M. Nießner, Scenecad: Predicting object alignments and layouts in rgb-d scans, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16, Springer, 2020, pp. 596–612.