Spatial Understanding for Industry

Daniele Di Mauro[1,*], Francesco Ragusa[1,*], Antonino Furnari[1] and Giovanni Maria Farinella[1]

[1]Next Vision s.r.l., Viale Andrea Doria, 6, Catania, Italy

Abstract

Spatial computing—the convergence of digital and physical space through technologies like AR, VR, IoT, and AI—is rapidly reshaping both industrial operations and healthcare delivery. This paper explores how spatial intelligence enables more intuitive, context-aware interaction with environments and data, unlocking new capabilities. We present two illustrative case studies. The first, SAKURA, focuses on a rehabilitation system that uses spatial computing and motion tracking to create immersive therapy environments for patients recovering motor function. This approach not only enhances engagement and adherence but also enables real-time feedback and remote monitoring by clinicians. The second case, NEPTUNE, examines the use of digital twins in underwater environments, where spatially intelligent systems integrate sensor data, 3D mapping, and simulation inspection in harsh and inaccessible conditions. Together, these cases demonstrate the transformative potential of spatial computing to bridge virtual and physical domains, offering scalable solutions for both human-centered care and complex industrial challenges.

Keywords

Wearable Devices, Industrial Setting, Digital Twin

1. Introduction

In the rapidly evolving technological landscape, spatial computing [1] represents a revolutionary paradigm shift in how humans interact with digital systems and their physical environments. Unlike traditional computing interfaces that separate users from data through screens and keyboards, spatial computing dissolves these boundaries by integrating digital capabilities directly into the physical world. This emerging field—representing the convergence of augmented reality (AR), virtual reality (VR), the Internet of Things (IoT), and artificial intelligence—creates experiences where digital and physical realities not only coexist but actively respond to and enhance one another. Spatial computing systems fundamentally transform user experiences by recognizing, interpreting, and adapting to physical spaces. They create digital overlays on physical environments, understand human movements and intentions within those spaces, and enable intuitive interactions that feel natural rather than mediated. This intelligent awareness of physical context represents a significant departure from traditional computing paradigms, enabling technology to respond not just to direct inputs but to spatial relationships, movements, and environmental factors. The implications of this shift extend far beyond consumer applications or gaming. In industrial settings, spatial computing enables workers to visualize complex information overlaid directly on physical equipment, access real-time data in context, and collaborate with remote experts who can see what they see. In healthcare, it creates new possibilities for patient care, rehabilitation, and medical training that were previously impossible to implement at scale. This paper examines how spatial intelligence—the core capability underlying spatial computing—enables more intuitive, context-aware interactions with environments and data, unlocking capabilities that address complex challenges. We present two illustrative case studies that demonstrate the transformative potential of these technologies in two very different scenarios: healthcare delivery and underwater operations. The first case study, SAKURA, explores an innovative rehabilitation system that leverages spatial computing and precise motion tracking to create immersive therapeutic

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
ddimauro@nextvisionlab.it (D. Di Mauro); fragusa@nextvisionlab.it (F. Ragusa); afurnari@nextvisionlab.it (A. Furnari); gmfarinella@nextvisionlab.it (G. M. Farinella)
0000-0002-4286-2050 (D. Di Mauro); 0000-0002-6368-1910 (F. Ragusa); 0000-0001-6911-0302 (A. Furnari); 0000-0002-6034-0432 (G. M. Farinella)

environments for patients recovering motor function. The system provides real-time feedback to patients, and generates comprehensive data on recovery trajectories. This spatial computing application demonstrates how digital interventions can meaningfully enhance human-centered care while expanding access to specialized rehabilitation services. The second case, NEPTUNE, study examines the deployment of digital twins in challenging underwater environments, where spatially intelligent systems integrate multiple data streams to create comprehensive virtual representations of submerged environment. These systems combine sensor networks, 3D mapping technologies, and simulation capabilities to enable inspection and maintenance planning for facilities that are physically hazardous or inaccessible to human workers. By creating accurate spatial models that evolve in real-time, these applications demonstrate how spatial computing can extend human capabilities in extreme environments while reducing operational risks. Together, these cases illustrate spatial computing's unique capacity to bridge virtual and physical domains, offering scalable solutions for both human-centered care and complex industrial challenges. As this technology continues to mature, it promises to fundamentally transform how we interact with our environments, access information, and solve problems across numerous domains and industries.

(a) Hololens 2[1]

(b) Meta Quest 3[2]

2. Sakura: System for Augmented Kinetics and Unified Rehabilitation Assessment

The SAKURA project (System for Augmented Kinetics and Unified Rehabilitation Assessment) involves the development of an intelligent system based on Computer Vision and Artificial Intelligence algorithms and augmented reality (AR) technologies for remote medical rehabilitation. The system allows the doctor to wear an augmented reality headset and perform exercises that involve interaction with objects (e.g., pouring a glass of water) that the patient, located in a remote location, wearing another augmented reality headset, will then repeat. The doctor's interaction with the real objects present in the environment is virtualized and displayed in the patient's environment as holograms, allowing the patient to perform the doctor's gestures with high precision. In order to create the SAKURA system, it is necessary to study and design artificial intelligence algorithms capable of automatically analyzing egocentric images and videos to infer the 3D position and shape of hands and objects with which users interact in the reference system of the environment in which the user is located. The discrepancy between the actual position of the hands and objects and the ideal position will be quantitatively measured, providing detailed feedback for follow-up and rehabilitation assessment.

2.1. Architecture

The HoloLens 2 device (see Figure 1a) is used for data capture and visualization, on both ends (user and doctor), streaming the user's field of view to the doctor supervising the rehabilitation task. The system

[2]https://learn.microsoft.com/en-us/hololens/

[2]https://www.meta.com/it/quest/quest-3/

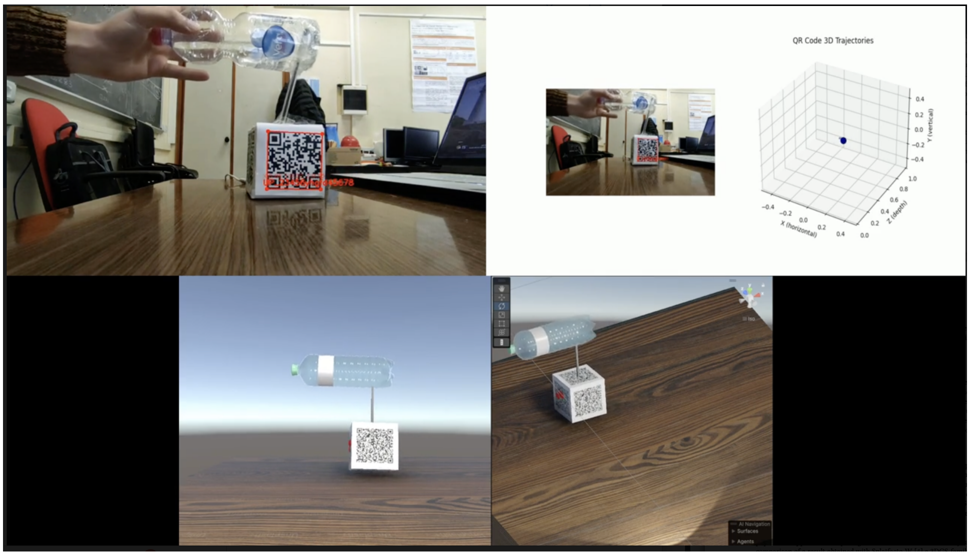

Figure 2: Multitracking and 3D simulation

utilizes an external GPU to compute the positions of the user's hands, the object being manipulated, and the exercise being performed. The HoloLens 2 offers reliable real-time recognition of audio commands and gestures. Throughout the rehabilitation process, the system collects exercise statistics to evaluate the user's progress.

2.2. State of The Art behind

3D hand and object pose estimation is a key technique with applications in AR, VR, and robotics. While most prior work has focused on single-hand poses [2, 3, 4] and single-hand object interactions [5, 6, 7], real-world scenarios often involve two hands manipulation. The H2O dataset [8] addresses this gap by providing annotations for both hands, object poses, and action classes. This project aims to develop and integrate advanced 3D pose estimation algorithms using real egocentric data, specifically within healthcare and rehabilitation settings.

3. Neptune: Neural Rendering & Edge AI Platform for 4D synthetic Twins generation during Underwater Navigation & Exploration

The project aims to design and implement a platform to manage optical sensors, edge ai onboard of underwater rovers and neural rendering models to create and certify digital twins of underwater ecosystems. Digital twins in this field are essential for exploratory, educational and informative purposes. The project which involves a pool of SMEs will be validated in two different contexts: the ecosystem of the Marine Protected Area of Portofino in Liguria and the Roman Port of Lipari in Sicily. A more detailed description of environments, 3D modeling used and wearable device application follows.

3.1. 2 Environments scenarios: Portofino and Lipari

The environments in the study represents two possible use case for underwater reconstruction: ecological maintenance and archeological surveys. The protected habitat of Portofino is defined in ecology as a coralligenous biocenosis, or coralligenous community, and is characteristic of depths greater than 30-35 m. It is an ecosystem that develops on a base of encrusting calcareous algae with two-dimensional growth along the rocky slope, on which the "animal forests" made up of gorgonians grow. Gorgonians have a three-dimensional calcareous skeleton that develops in a fan shape, arranging itself perpendicularly to the current to which it is subject. The species of gorgonians in the areas of the Portofino

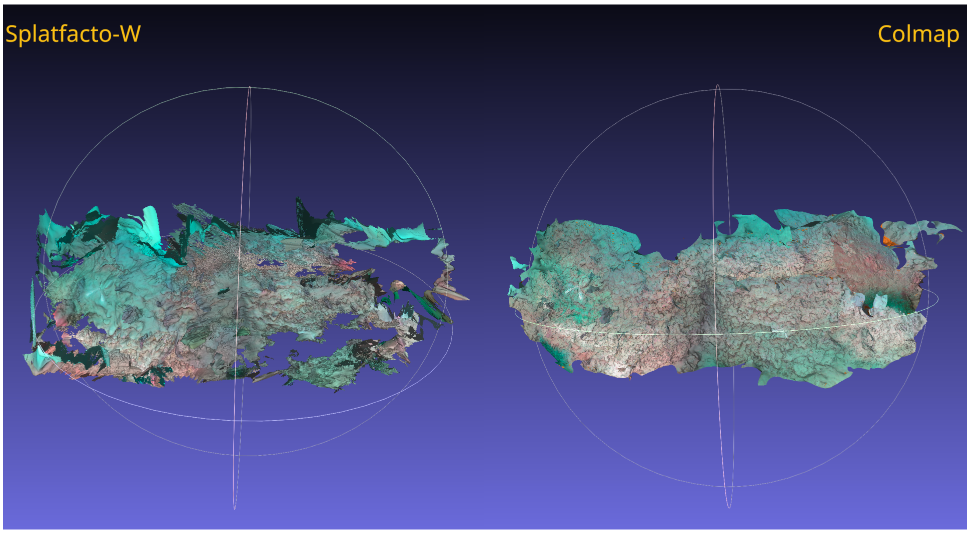

Figure 3: Comparison of a 3D Gaussian Splatting mesh reconstruction (Splatfacto-W) vs Structure From Motion (Colmap)

promontory are mostly *Paramuricea Clavata* (red gorgonian) which does not exceed 50-60 cm in height, and *Eunicella Cavolinii* (yellow gorgonian), around 20-30 cm. The skeleton is hard, being made of limestone, but with a good degree of flexibility if subjected to light impacts.

The archaeological area covered is the one near the wreck of the Capistello shoal. The wreck is the only Aeolian wreck whose wooden structure has also been examined. The cargo, dating back to the beginning of the 3rd century BC, consists of Greek-Italic amphorae and black-glazed ceramics of Campanian production. The ship probably came from one of the Greek centers in the Bay of Naples, to head, via the Aeolian Islands, to one of the Punic centers of western Sicily, such as Lilybaeum (Marsala), and towards North Africa (Carthage).

3.2. Neural Modeling

The work focuses on evaluating the maturity of neural modeling methods that have emerged in the last five years, specifically Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS).
Neural Radiance Fields (NeRF) [9] was a seminal work for 3D scene reconstruction and novel view synthesis using neural networks. By learning a continuous volumetric scene function from a sparse set of 2D images, NeRF can generate photorealistic renderings from any viewpoint without explicitly modeling geometry. This technique works by encoding how light travels through and interacts with space. Since its introduction in 2020, NeRF has sparked a revolution in computer vision, with applications ranging from virtual reality and film production to architecture visualization and digital heritage preservation.
3D Gaussian Splatting, introduced in [10], improve on NeRF using explicit 3D Gaussian primitives rather than NeRF's implicit neural fields. While NeRF encodes scene information within neural network weights and requires computationally expensive ray marching during rendering, Gaussian Splatting explicitly models the scene with oriented 3D Gaussians that can be directly rasterized using traditional graphics pipelines, enabling real-time frame rates (>100 FPS). In Figure 3 a comparison of a result obtained with Splatfacto-W [11] a 3DGS-family algorithm, and Colmap, a state of art SfM application.

3.3. Wereable Device Application

The developed wearable application allows to navigate 3D reconstructions in different modalities. The user experiences a fully immersive underwater reality with interactive 3D model exploration. Through controllers, the user can move freely in the environment simulating the underwater drone movements.

The current implementation uses Meta Quest 3 (see Figure 1b). The analog sticks will allow you to move in space and height, while the right and left triggers will allow you to rotate the view. Each movement is accompanied by a sound effect to simulate the underwater movement of the drone. Using the stick, you can activate or deactivate a virtual torch that, thanks to the lighting and a particle system, simulates an underwater torch, allowing you to illuminate the surfaces of the 3D model in detail. The torch is also accompanied by a fog effect generated by particles, to make the disturbances of the light beam under the surface of the water more realistic.

The system is equipped with a UI designed to provide additional information (when available) about the environment visited. This information can include pressure, temperature, oxygen and water quality and is displayed via an interface that follows the left controller. Since the 3D model is based on acquired data, the information is spatially discretized.

4. Conclusion

This work presented two projects that apply spatial intelligence to telemedicine and underwater monitoring. These cases demonstrate the transformative potential of spatial computing to bridge virtual and physical domains. The projects highlight the critical role of wearable devices and ego-centric vision in leveraging digital twin technologies to create more effective and responsive systems across diverse application domains.

Acknowledgments

The research described in this paper is supported by the project NEPTUNE - RAISE - PNRR Missione 4 Componente 2 - Dalla ricerca all'impresa - INVESTIMENTO 1.5 - Creazione e rafforzamento di Ecosistemi dell'innovazione costruzione di leader Territoriali di R&S - Bando a cascata 2024 - CUP J53D23020140005, and by the project SAKURA - PNRR Missione 4 Componente 2 - Dalla ricerca all'impresa - Ecosistemi dell'Innovazione iNEST - Interconnected Nord-Est Innovation Ecosystem Spoke 2 - Health, Food & Lifestyle - Bando a cascata 2024 - Second Call, CUP E63C22001030007.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. Shekhar, S. K. Feiner, W. G. Aref, Spatial computing, Communications of the ACM 59 (2015) 72–81.

[2] S. Baek, K. I. Kim, T.-K. Kim, Augmented skeleton space transfer for depth-based hand pose estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8330–8339.

[3] S. Baek, K. I. Kim, T.-K. Kim, Pushing the envelope for rgb-based dense 3d hand pose estimation via neural rendering, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 1067–1076.

[4] D. U. Kim, K. I. Kim, S. Baek, End-to-end detection and pose estimation of two interacting hands, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11189–11198.

[5] S. Baek, K. I. Kim, T.-K. Kim, Weakly-supervised domain adaptation via gan and mesh model for estimating 3d hand poses interacting objects, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 6121–6131.

[6] Y.-W. Chao, W. Yang, Y. Xiang, P. Molchanov, A. Handa, J. Tremblay, Y. S. Narang, K. Van Wyk, U. Iqbal, S. Birchfield, et al., Dexycb: A benchmark for capturing hand grasping of objects, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 9044–9053.

[7] J. Park, Y. Oh, G. Moon, H. Choi, K. M. Lee, Handoccnet: Occlusion-robust 3d hand mesh estimation network, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 1496–1505.

[8] T. Kwon, B. Tekin, J. Stühmer, F. Bogo, M. Pollefeys, H2o: Two hands manipulating objects for first person interaction recognition, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10138–10148.

[9] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, Nerf: Representing scenes as neural radiance fields for view synthesis, Communications of the ACM 65 (2021) 99–106.

[10] B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3d gaussian splatting for real-time radiance field rendering, ACM Transactions on Graphics 42 (2023). URL: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

[11] C. Xu, J. Kerr, A. Kanazawa, Splatfacto-w: A nerfstudio implementation of gaussian splatting for unconstrained photo collections, 2024. URL: https://arxiv.org/abs/2407.12306. arXiv:2407.12306.