Project - Intelligent Digital Surgeon (IDS)

Number: 100.133 IP-ICT

For reference and additional details, this document is an extended version of the summarized final status report.

Abstract

The Intelligent Digital Surgeon (IDS) project[1], a collaboration between MIRALab SARL (MIRALab), the University of Geneva (UNIGE), the Hôpitaux Universitaires de Genève (HUG), and ORamaVR S.A (ORamaVR), addressed a significant gap in medical education: the limited availability of specialized, real-time feedback during surgical training.

In this context, the IDS project developed a system for real-time surgical gesture recognition and immediate personalized feedback. Leveraging VR, Machine Learning (ML), and Large Language Models (LLMs), we have collected, analyzed, and evaluated surgical gestures, offering trainees real-time, expert-aligned feedback based on curated medical knowledge. Concretely, achieving this objective included creating an open dataset of professional and trainee gestures (to be finalized as soon as we receive the approval from the ethical committee for the collection of students' gestures), developing a state-of-the-art gesture recognition algorithm, and implementing an animated digital surgeon that provided 3D visualizations and natural language explanations of errors. In summary, we have worked as planned:

1. Analysis of the surgeons' needs and their gestures;
2. The design and implementation of AI algorithms to recognize gestures;
3. The design and implementation of AI algorithms to compare the students' gestures with the dataset containing the surgeons' gestures;
4. A first interface to communicate with the Virtual AI surgeon.

***Keywords**:* Virtual Reality, Gesture Analysis, Surgical Training, Intelligent Surgeon, 3D Modeling, Animated Intelligence, User Needs.

[1] The project website can be accessed at https://intelligentdigitalsurgeon.miralab.com/.

Question 1.1 - Summarize the progress of the activities in relation to the project planning and indicate to which degree the project objectives have been met.

The Intelligent Digital Surgeon (IDS) project successfully has met its primary objective: developing an intelligent, real-time feedback system for surgical training within the MAGES Virtual Reality (VR) platform. The IDS addressed a critical gap in providing accessible, expert-level feedback in surgical training. By integrating deep learning for gesture recognition, 3D visual feedback, and a Large Language Model (LLM) for interaction between the trainee and the system, the IDS presented an interesting learning experience with real-time surgical training support.

Next, we describe the five specific work packages (WPs) to achieve our goals: *1) WP1 Use case, Requirements, and Architecture Specification; 2) WP2 Design and Implementation of Intelligent Digital Surgeon (IDS); 3) WP3 Implementation of Edge Computing Scheme for VR Supported by 5G Networks; 4) WP4 MAGES SDK physics-based cutting & tearing engine based on Geometric Algebra; and 5) WP5 System Integration, Training module development and evaluation*.

WP1: Use case, Requirements, and Architecture Specification

This WP was completed within the designated timeline, identifying essential user requirements, use case scenarios, and evaluation criteria. The deliverable consisted of a comprehensive report defining optimal surgical gestures and architecture specifications. Specifically, this WP comprised two tasks, as described next.

Task 1.1: Definition and Analysis of Requirements, Use Case Scenarios, Evaluation criteria (ORAMA, HUG, MIRALab SARL, UNIGE)

Task Status: Completed

This task aimed to create a robust and reliable training system. To accomplish that, we defined two types of gestures, 1) incision; and 2) suturing. Each gesture comprises different steps, as described in Tables 1 and 2, covering the whole surgical procedure.

Step	Description	Step	Description
I1	Grasping scalpels or surgical forceps	S1	Grasping a forceps and a needle holder
I2	Maneuvering towards the targeted tissue site	S2	Securing the needle within the needle holder
I3	Securing the surrounding skin	S3	Maneuvering towards the targeted tissue site
I4	Executing a skin incision using the scalpel	S4	Elevating one edge of the skin
I5	Mobilizing the incised skin with the scalpel	S5	Guiding the needle through the lifted layer
		S6	Drawing the suture through the tissue
		S7	Transitioning between instruments
		S8	Trimming the excess suture material
		S9	Executing a secure knot
		S10	Re-grasping the needle holder

Table - Steps for performing an incision and suturing

- **Deliverables**
    - D1.1: User Requirements, Use Case Scenarios, and Evaluation Specifications Report - This report documented the outcomes of Task 1.1, detailing the user

requirements, specific use cases, and the evaluation criteria for assessing the IDS system's performance in training environments.

- **Milestones**
    - MS1: Requirements, Architecture and Evaluation Specifications
        - Goal 1: User requirements (D1.1) - Completed
        - Goal 2: Use case scenarios (D1.1) - Completed
        - Goal 3: Evaluation criteria (D1.1) - Completed

Task 1.2: Software Architecture Specification

Task Status: Completed

This task defined the software architecture to meet the requirements from Task 1.1. Specifically, we have defined two main components: 1) the Distributed Application Architecture; and 2) the Component and Interface Specifications.

Distributed Application Architecture. To accomplish our vision within the IDS context, we optimized the MAGES SDK, which is built on top of a distributed application architecture designed to enhance the performance and scalability of IDS' VR training system. This architecture separates computation-intensive tasks and distributes them across multiple servers and services, allowing MAGES to maintain high-performance VR interactions even in resource-constrained environments.

Concretely, MAGES balances high-complexity operations with real-time responsiveness, ensuring that our IDS system operates efficiently within VR environments through this distributed architecture.

Component and Interface Specifications: Detailing interactions between IDS modules and the partners' roles, including interfaces for real-time gesture recognition, 3D feedback, and data management. The development of the IDS system involved integrating gesture recognition, medical knowledge, and 3D visualization, ensuring the accuracy of both the movements and the feedback provided. Figure 1 illustrates how the different components of our solution interacted. It is possible to see that the workflow was divided into three main parts: AI Development, Application Development, and Evaluation. Next, we describe each part and its contribution to the overall development of IDS.

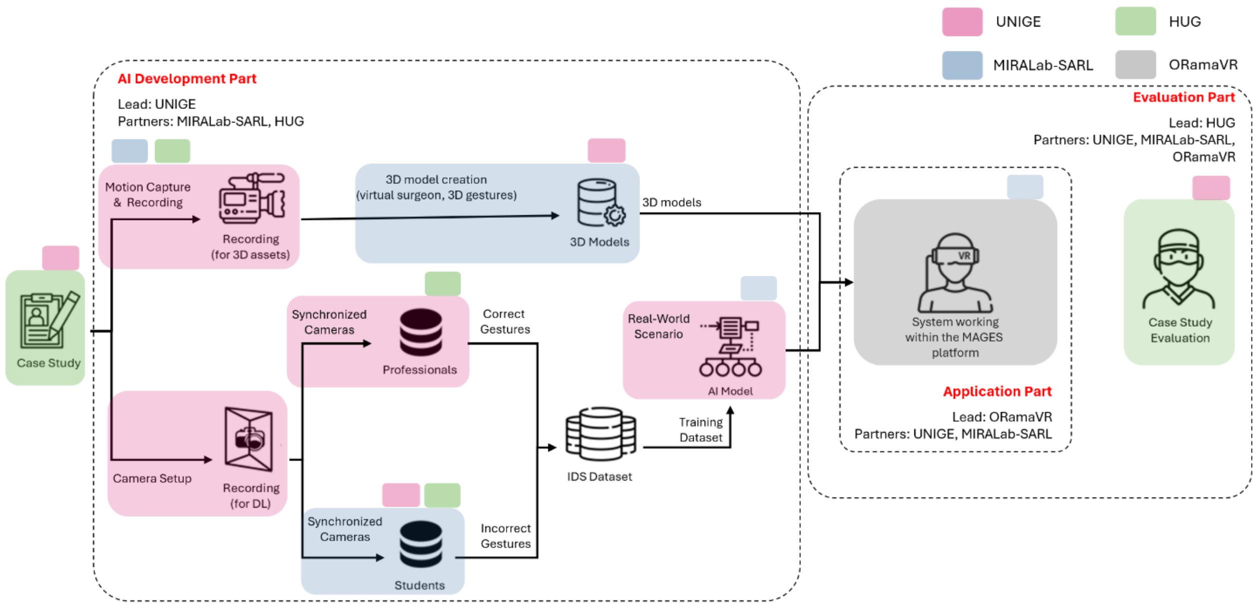

Figure 1- This architecture describes the workflow of the Intelligent Digital Surgeon (IDS) system, divided into three main parts: AI Development Part, Application Development Part, and Evaluation Part. Each part contains distinct components that interact to facilitate the collection, processing, and application of data for virtual surgical training. Each of these components is represented by color-coded boxes in the diagram. Pink represents data collection and AI development components (UNIGE), Blue highlights data storage, model creation components, and application settings for the real-world scenario (MIRALab), Green designates case studies and evaluation processes (HUG), and Gray indicates the application settings for VR scenarios (OramaVR).

AI Development: This involved data collection and model creation, professional recordings, and 3D model generation.

- Case Study: Initiated the process by defining the requirements and use cases that guide data collection and model design.
- Motion Capture & Recording: Professional surgeons were recorded using motion capture technology to generate data for 3D assets, focusing on accurate capture of surgical gestures.
- Camera Setup for Deep Learning (DL): This component involved setting up synchronized cameras to capture gestures from different angles. It allowed the recording of surgical actions by both professionals and students, with the recordings organized for deep-learning training.
- Recordings (for DL): Two datasets were generated from synchronized camera recordings:

    - Professionals: Data collected from professional surgeons to create a gold standard for surgical gestures.
    - Students: Data collected from medical students, providing a range of skill levels and common errors for training the AI model (Partially Completed).
- 3D Model Creation: Using recorded data, 3D models were generated. These models represented a virtual surgeon and various surgical gestures, forming the basis for the animated feedback component in the IDS.
- AI Model: Based on the collected data from professionals and students, an AI model was trained to recognize surgical gestures. This model was the basis for creating the gesture recognition module in the VR environment.

Application Development: This involved the implementation of the IDS in two different environments: real-world scenarios (the baseline) and VR environment (the proposed solution).

- Real-World Scenario: Before creating the VR solution, we developed a baseline system that could operate using a camera and computer setup, testing the gesture recognition module in a non-VR environment. This setup was useful to assess different deep learning techniques for gesture analysis before applying it to a VR environment, providing a tool for validation during the development process.
- VR Scenario: The IDS system operated within a VR environment, providing an immersive training experience. In this scenario, the AI model detects gestures and provides feedback within the VR system, simulating a realistic surgical training environment.

Evaluation: This included the validation of the IDS system with real users in the VR environment (Partially Completed).

- Case Study Evaluation: The system's effectiveness was evaluated through studies involving medical students and professionals. This evaluation process collected feedback on the system's accuracy, responsiveness, and overall educational value (Partially Completed).

In summary, this workflow outlines the data flow from initial recording and data collection, through AI model training, to the application of the IDS system in real-world settings as baseline and VR settings as our solution. The evaluation phase completed the cycle, ensuring IDS met its objectives for real-time feedback in surgical training.

Deliverables

- D1.2: Software Architecture Specification Report - In the MAGES' platform paper[2], we described the architecture and presented the innovations of our approach.

Milestones

- MS1: Requirements, Architecture and Evaluation Specifications
    - Goal 4: Application architecture (D1.2) - Completed

WP2: Design and Implementation of Intelligent Digital Surgeon (IDS)

WP2 focused on the practical development of the Intelligent Digital Surgeon (IDS), specifically on creating the essential components of our solution: 1) the animated digital surgeon; 2) the machine learning-based gesture analysis module; and 3) the recommendation module.

Task 2.1: Creation and animation of a Cognitive and Intelligent Digital Surgeon (UNIGE, MIRALab SARL)

Task Status: Completed

This task aimed to develop and animate the IDS, which functioned as a cognitive agent within the MAGES VR platform. A video demonstration can be accessed at: https://www.youtube.com/watch?v=eK8yAozEN6o. The animated IDS served as an interactive, intelligent instructor, guiding trainees through surgical procedures with real-time visual and verbal feedback. Since we focused on building a realistic, anatomically accurate digital surgeon, Figures 2 and 3 show examples of how we used data collected from professional surgeons (as described in Task 2.2) for incision and suturing procedures, respectively. The animated IDS accurately represented key hand and instrument movements required for surgical procedures. The animated IDS accurately represented key hand and instrument movements required in surgical tasks and was further enhanced with animations displaying proper techniques from different angles, along with common corrective actions to help trainees visualize ideal procedural steps. A notable feature of this animated IDS, achieved through detailed modeling, was its ability to dynamically adjust its demonstration in real time based on identified mistakes (based on the steps described in Tables 1 and 2). For example, if the system detected improper needle handling, the 3D demonstration would specifically focus on the correct method for grasping and guiding the needle.

[2] Zikas, Paul, et al. "MAGES 4.0: Accelerating the world's transition to VR training and democratizing the authoring of the medical metaverse." IEEE Computer Graphics and Applications 43.2 (2023): 43-56.

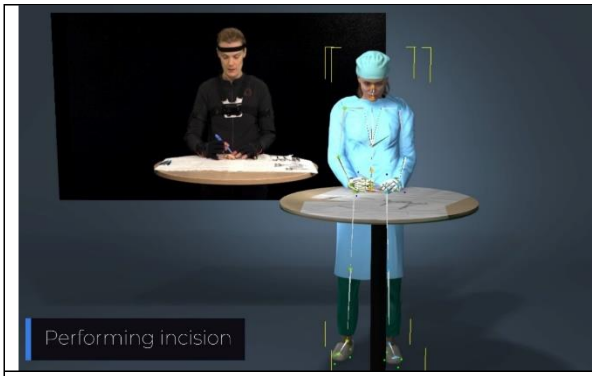

Figure 2 - A professional surgeon performing an incision at the UNIGE recording studio (motion-capturing technology), accompanied by a 3D model with skeleton annotation replicating the same incision gesture. For each gesture, the surgeon explained the action to be performed, facilitating annotation of the recording data.

Figure 3 - A professional surgeon performing suturing at the UNIGE recording studio (motion-capturing technology), accompanied by a 3D model with skeleton annotation replicating the same suturing gesture. For each gesture, the surgeon explained the action to be performed, facilitating annotation of the recording data.

Figure 4- All-in-one motion capture solution using the Smartsuit Pro II and Smartgloves. The whole-body movements were recorded, including the hands' gestures. All components are portable, allowing the expert to perform gestures as they would in a real medical environment.

The motion-capturing process, using the Smartsuit Pro II and Smartgloves (Figure 4), involved recording an expert surgeon performing surgical gestures in a controlled environment (UNIGE studio). This step was essential to ensure that the demonstrated gestures were anatomically correct and accurately reflected real-world surgical techniques. Specifically, we employed 3D rendering techniques to create models of the hands and instruments, illustrating their interactions with surgical tools and tissues. The completed setup, which enabled us to capture these intricate movements from multiple angles and perspectives, is illustrated in Figure 5.

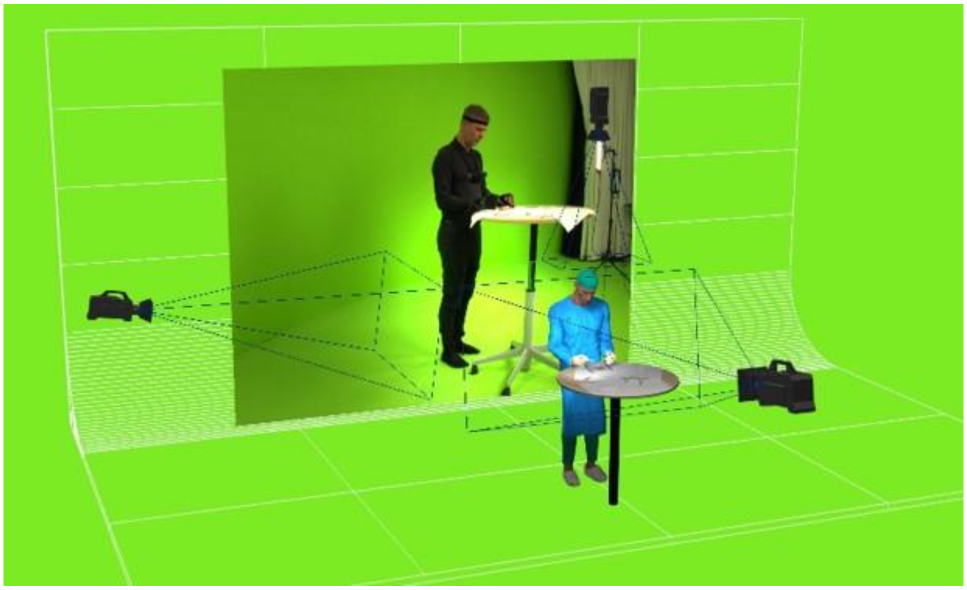

Figure 5 - Recording setup at the UNIGE studio, showing a professional in a controlled environment with multiple cameras positioned at different angles and focal lengths. This multi-angle configuration enables comprehensive 3D model alignment, essential for animating the digital surgeon with precise and anatomically accurate movements.

- **Deliverables**
  - D2.1: Animated Intelligent Digital Surgeon (UNIGE, MIRALab SARL, HUG, M12) - The platform with the animated digital surgeon can be accessed at https://intelligentdigitalsurgeon.miralab.com/

Task 2.2: Analysis of the trainee's hand/arm gestures and recommendation (UNIGE, MIRALab SARL, HUG)

Task Status: Partially Completed

This task involved collecting a comprehensive dataset of surgical gestures performed by both professional surgeons and medical trainees to train the gesture analysis model. The data collection process focused on capturing real-life representations of surgical gestures, ensuring the model's ability to recognize and evaluate skill levels and common errors. The dataset included:

- Dataset of Professional Surgeon Gestures: This dataset included the techniques performed by professional surgeons for the incision and suturing procedures. Serving as a gold standard for training and evaluation, it includes data from a panel of 20 professional surgeons from HUG, considering all the steps outlined in Tables 1 and 2. Each surgeon performs these tasks naturally, guided only by a brief description of the expected technique, with each gesture repeated three times. Incision and suturing were selected as fundamental skills required by all surgeons. To simulate realistic conditions, a chicken thigh was used as a surgical model, chosen for its accessibility and structural similarity to human skin. Surgeon performances were recorded using fixed cameras from multiple viewpoints (Figure 6), making the system robust to perspective variations encountered in real practice and ensuring consistent gesture recognition across different angles.

- Dataset of Trainee Gestures (To be completed): This dataset gathers data from medical students performing similar tasks, capturing a range of skill levels and deviations from the standard. The variation in skill levels provides the IDS with the necessary data to recognize both correct and incorrect gestures. Approval from the UNIGE ethical committee is required to complete this task, and we are in the process of securing this authorization to begin data collection. Based on Deliverable D1.1 and the Objective Structured Assessment of Technical Skills (OSATS) for surgical skills, we identified various common errors students make (e.g., respect for tissue, time and motion, instrument handling, flow of operation), which will be labeled for validation. In this project, the OSATS tool has been extended to provide more detailed feedback. Thus, rather than only assigning scores on a Likert scale for each criterion, our tool provides targeted descriptions of mistakes. For instance, rather than simply indicating an "Instrument handling mistake," feedback might specify, "Issue with instrument handling: the needle was not properly secured within the holder." The list below presents the details for each of the error types, illustrating how OSATS was adapted for IDS requirements.
- Dataset of Common Mistakes and Corrective Suggestions (To be completed): Using data from the first two datasets, this third dataset focuses on the most common types of mistakes identified, accompanied by corrective suggestions. It includes 3D models that will visually demonstrate the correct techniques, helping trainees understand and correct their gestures effectively.

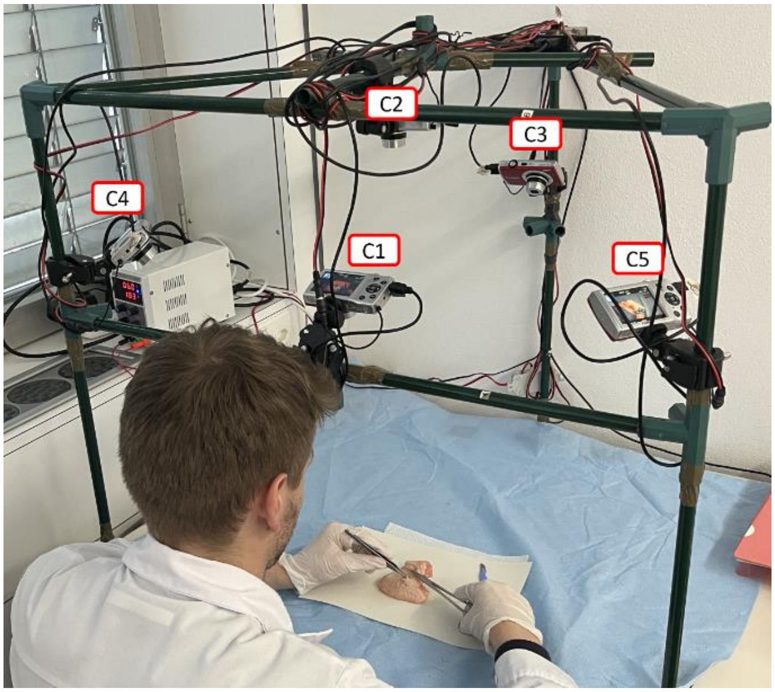

Figure 6 - Equipment setup for capturing gesture data with five cameras (C1 to C5), each positioned at different angles to provide comprehensive views for gesture analysis and identification. In this example, a professional demonstrates a surgical gesture on a chicken leg, simulating real-world conditions.

Respect for Tissue	- Incorrectly elevating one edge of the skin - Incorrectly securing the surrounding skin - Incorrectly mobilizing the incised skin with the scalpel - Excessive force on tissues - Poor stabilization during needle insertion
Time and Motion	- Incorrect suturing time - Incorrect incision time
Instrument Handling	- Incorrectly grasping a forceps and a needle holder - Incorrectly securing the needle within the needle holder - Incorrectly maneuvering towards the targeted tissue site - Incorrectly guiding the needle through the lifted layer

	- Incorrectly drawing the suture through the tissue - Incorrectly transitioning between instruments - Incorrectly trimming the excess suture material - Incorrectly executing a secure knot - Incorrectly re-grasping the needle holder - Incorrectly grasping scalpels or surgical forceps - Incorrectly executing a skin incision using the scalpel - Dropping tools
Flow of Operation	- Wrong sequence of actions

After collecting the dataset, we developed an algorithm to serve as a baseline gold standard in real-world settings. This algorithm was capable of recognizing surgical gestures and specific procedural steps from videos recorded with 2D cameras. It employed a spatial-temporal transformer-based model designed for surgical gesture recognition.

The model processes these inputs as continuous transformation data, effectively including position, rotation, and movement trajectories of both hands and surgical instruments. The gesture recognition model, now optimized for MAGES' VR, uses these data points to evaluate the accuracy, precision, and fluidity of each gesture. We trained a modified deep learning model to assess key phases within each surgical task, aligning gestures with predefined labels based on expert input (e.g., poor, mediocre, excellent). Specifically, we created a 15-layer CNN architecture to efficiently process trajectory data and apply augmentation techniques to reduce the need for extensive labeling. With this setup, the VR environment system can capture complex, frame-by-frame movement details and provide immediate feedback on specific aspects of the technique. This integration allows the IDS system to assess real-time performance metrics like tool handling and motion trajectory, and provide additional improvement recommendations, offering trainees targeted, expert-aligned guidance in an immersive virtual environment.

- Deliverables
    - D2.2: Creation of datasets (UNIGE, MIRALab SARL, HUG, M2) - We have completed the creation of the dataset with professional surgeons. Although

we have not collected data from medical student trainees, we have established the data collection protocol and are waiting for the approval of the UNIGE ethical committee to start collecting this data. We will then make the datasets open source.

  - D2.3: Recognition and Analysis of the trainee's hand/arm gestures (UNIGE, MIRALab SARL, M13) - We have developed the algorithm for the gesture analysis module directly from video data used as the baseline and the gesture recognition on VR environments within the MAGE platform.

- Related Publications:
  - TransSG: A Spatial-Temporal Transformer for Surgical Gesture Recognition
  - MAGES 4.0: Accelerating the World's Transition to VR Training and Democratizing the Authoring of the Medical Metaverse

Task 2.3 Recommendation by the digital surgeon to the trainee (UNIGE, MIRALab SARL, HUG)

Task Status: To be completed soon

The IDS integrates Large Language Models (LLMs) to provide verbal explanations alongside visual feedback. For example, when IDS detects an error, such as improper needle handling during suturing, it visually demonstrates the correct technique through a 3D animated model (Figure 7) and provides a verbal explanation that clarifies the specific mistake and emphasizes the correct approach (Figure 8). This combination of real-time visual guidance and explanatory dialogue aids trainees in grasping complex movements and understanding corrective actions, improving their practical skills.

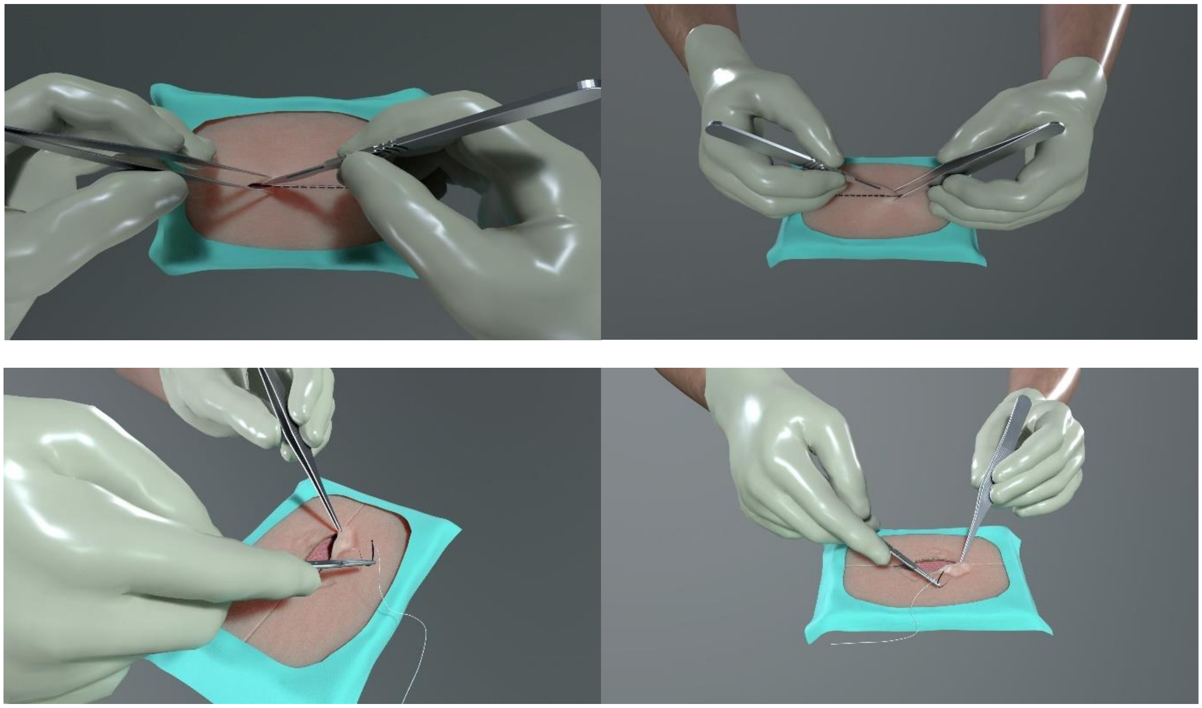

Figure 7 - Examples of different views that can be presented to medical trainees observing the explanation of a gesture being performed by a professional. In these figures, we aim to highlight how presenting different angles of the same gesture conveys different information, and different people may benefit from their preferred viewpoint for learning. In this project, the different views can also be used by expert evaluators to better understand the students' gestures.

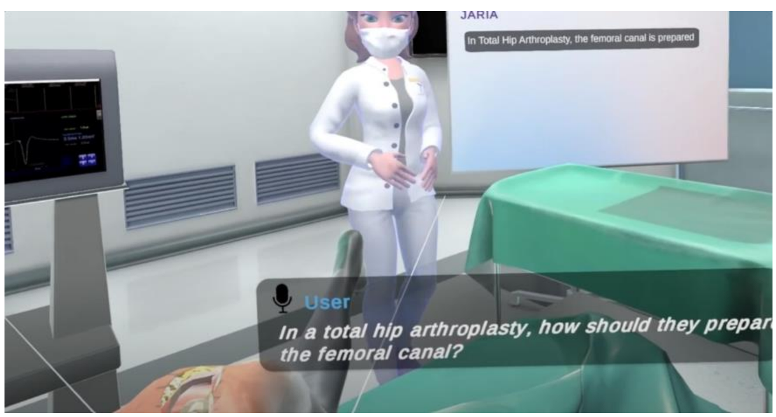

Figure 8 - Interaction within a VR environment between the user and the digital surgeon, where the user inquires about the preparation of the femoral canal in a total hip arthroplasty procedure. The digital surgeon responds with relevant guidance, simulating a real-time consultation. This setup illustrates the educational potential of the animated intelligent surgeon in providing interactive, procedural advice in response to user queries, enhancing medical training through immersive VR experiences.

Specifically, to create this interactive feature, we integrated an LLM with a Retrieval-Augmented Generation (RAG) framework.

With RAG, the animated digital surgeon can dynamically pull relevant content from our medical knowledge base whenever a trainee makes a query or when the system detects a mistake. For instance, if a medical student performs a suturing step incorrectly, the RAG-enabled LLM can retrieve specific information related to needle handling from the repository (as illustrated in the data flow diagram in Figure 9). This information is then combined with the LLM's generative abilities to produce a well-informed and contextually accurate explanation, which is delivered verbally alongside the visual demonstration of the correct technique.

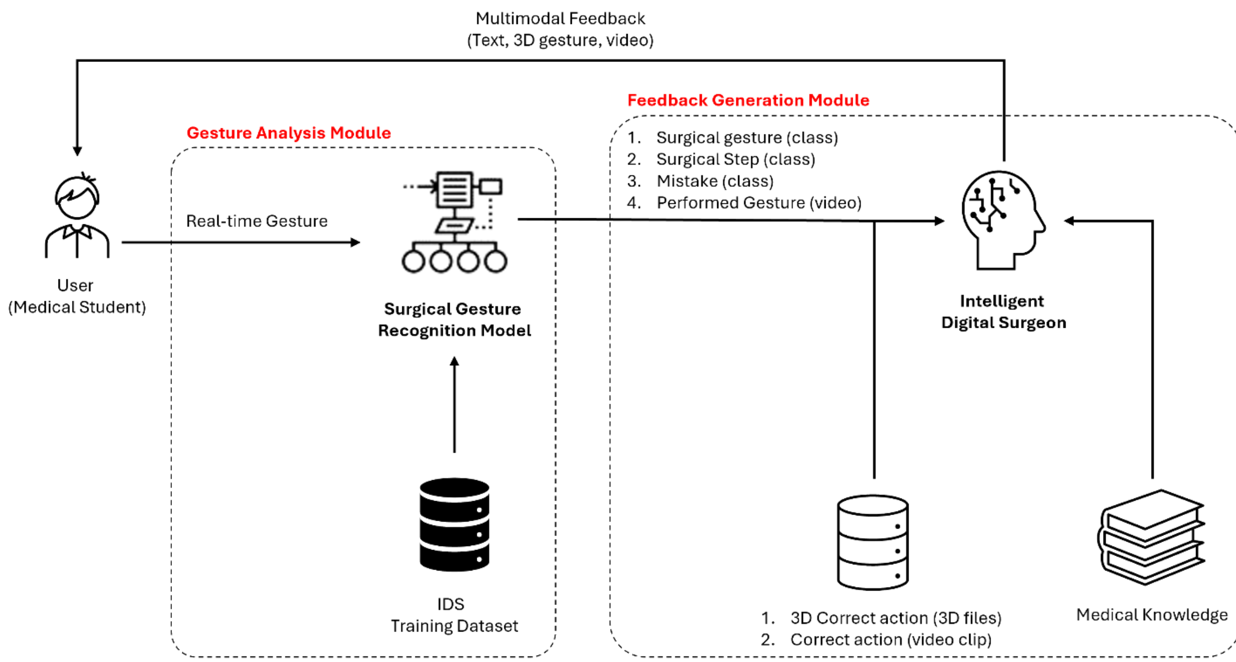

Figure 9- This diagram illustrates how data flows through the recommendation system, where user interactions are processed, relevant information is retrieved from the medical knowledge base, and recommendations are generated in real-time to support surgical training and feedback. To integrate this knowledge, we used a multi-modal LLM with RAG.

- **Deliverables**
  - D2.4: Recommendation to the trainee by the IDS (UNIGE, MIRALab SARL, HUG, M15) - A demo of this application can be accessed here.
- **Milestones**
  - MS4: Components-microservices Implementation
    - Goal 8: IDS components implemented and tested (D2.4) - Partially Completed, the dataset with students' gestures was not fully integrated and tested. This will be achieved within the next steps of our project, after obtaining ethical approval to conduct the data collection.

WP3 Implementation of Edge Computing Scheme for VR Supported by 5G Networks

Task 3.1: Design and implementation of edge server architecture (UNIGE, ORAMA)

Task Status: Completed

The architecture supporting ORamaVR's collaborative VR training applications is designed to meet the demanding requirements of real-time interactivity, scalability, and cross-platform compatibility. By leveraging edge computing resources and advanced networking technologies, the system ensures a seamless and immersive experience for users of untethered HMDs.

Key Components and Their Roles

At the heart of the system is a distributed architecture that incorporates mobile clients, edge services, and networking layers. These components work in unison to provide a responsive VR training environment:

1. **HMD Device (Mobile Client):** The HMD serves as the primary interface for users, handling essential UI elements and input/output operations. It streams user inputs—such as head movements, gestures, and interactions—directly to the Local Service (LS). In return, it receives encoded video streams for rendering, ensuring minimal computation overhead on the device itself. This lightweight approach enhances mobility while maintaining high performance.
2. **Local Service (LS):** Positioned at the edge of the network, the LS is the computational backbone of the system. By running the Unity Engine and MAGES SDK, it processes VR training content, handles scene rendering, and manages session data. Importantly, the LS operates independently of proprietary APIs, ensuring compatibility across different HMD platforms. Its ability to support multiple users simultaneously by initiating separate rendering and encoding instances makes it both scalable and efficient.
3. **Signaling Server:** Facilitating communication setup between the HMD and LS, the signaling server acts as the initial bridge for establishing WebRTC connections. Built as a lightweight Node.js application, it ensures seamless coordination of data channels and streaming protocols. This component is crucial for initiating direct, low-latency connections between devices.
4. **Relay Server:** In collaborative VR training scenarios, the relay server plays a pivotal role in managing multi-user interactions. It coordinates peer-to-peer communication, supports host migration to ensure continuity, and minimizes dependencies on the master server. By hosting this functionality on the edge, the architecture achieves faster synchronization and reduced latency for multi-user environments.

Technological Underpinnings

The architecture's success hinges on its use of state-of-the-art technologies and frameworks:

- **UDP Protocol:** Ensures efficient, low-latency data exchange, especially critical for real-time VR experiences.
- **Photon Relay Server:** Provides advanced networking features like host migration, message broadcasting, and state synchronization, enhancing collaborative functionality.
- **WebRTC for Unity:** Enables high-quality, real-time video and audio streaming between devices, ensuring smooth interactions even in complex network environments.
- **VM Packaging and Deployment:** LS and relay server instances are packaged as lightweight VM images, allowing seamless deployment across edge nodes with minimal configuration.

Seamless Integration and Performance

This architecture seamlessly integrates the capabilities of cloud and edge computing with robust networking solutions. By offloading heavy computations to edge nodes, the system ensures that mobile HMDs remain lightweight and responsive, while edge-based processing minimizes latency. Additionally, the choice of standard protocols and APIs guarantees cross-platform compatibility and scalability, making the system adaptable to future developments.

Future Enhancements for Sustained Excellence

While the architecture already provides a strong foundation for immersive VR training, ongoing enhancements will further refine the system:

- **Optimized Edge Node Placement:** Dynamic allocation of resources based on user proximity will further reduce latency and improve session responsiveness.
- **Adaptive Networking:** Real-time adjustments to data transmission rates based on network conditions will ensure a consistent Quality of Experience (QoE).
- **Enhanced Resource Management:** Automated provisioning of edge and cloud resources will simplify scaling and improve cost-efficiency.

By integrating these components and processes into a cohesive architecture, ORamaVR's platform provides a cutting-edge solution for collaborative VR training. The thoughtful balance of edge, cloud, and networking ensures a high-quality, scalable experience for untethered HMD users.

- **Deliverables**
    - D3.1: Edge server architecture report (UNIGE, ORAMA, M6) - This report provided the final architecture of the dedicated server, including a description of the underlying components and their interaction.
- **Milestones**
    - MS2: Edge server Implementation
        - Goal 5: Functional local edge server (D3.1) - Completed

Task 3.2: Deployment of edge server within the 5G edge-cloud continuum (UNIGE)

Task Status: Completed

The networking architecture and performance evaluations underpin collaborative VR applications deployed via the IDS platform. It defines workflows, experiments, and technical benchmarks across multiple scenarios, emphasizing performance optimization and user experience. The following sections delve into the core aspects of the system, from scenario descriptions to experimental findings, with a focus on measurable outcomes and future improvements.

Primary Scenarios and Their Workflows

The networking scheme supports three primary scenarios, designed to ensure seamless deployment and operation of collaborative VR applications:

Scenario 0: Application Submission

This scenario facilitates the application developer's interaction with the IDS platform to prepare for deployment.

- **Workflow:**
    - The developer accesses the IDS Platform portal.
    - Application-specific data is submitted to enable subsequent deployment.

Scenario 1.1: Starting the Application

This scenario governs the initiation of the application from the HMD and its deployment on the Local Service (LS).

- **Workflow:**

- The user sends a launch command from the HMD to the IDS platform via the event bus.
- The IDS platform deploys the application image on the LS.
- Streaming relies on **WebRTC for Unity**, which uses a third-party signaling server to establish peer-to-peer connectivity.

- **Details on WebRTC Implementation:**

WebRTC employs a signaling solution with a third-party server for connection establishment. Although counterintuitive for peer-to-peer setups, this method:

- Ensures robust connectivity in complex scenarios (e.g., NAT or firewall restrictions).
- Simplifies the discovery of direct routes between peers.

Scenario 1.2: Session Creation

This scenario enables users to create or join collaborative VR sessions.

- **Workflow:**
  - The user interacts with the HMD interface to access the session list.
  - A request for the relay server address is sent to the IDS platform.
  - The platform deploys a Photon relay server instance on an available cloud resource.
  - Once the relay server connection is established, the session list is retrieved, and a new session can be created for collaboration.

Key Prerequisites for All Scenarios:

1. Connectivity: Reliable communication via 5G or Wi-Fi between HMD and platform.
2. Event Bus Integration: The publish-subscribe model allows application components to send and receive messages efficiently.
3. App-to-App Communication: Direct app communication is managed using the UDP protocol for fast and lightweight data transfer.

Performance Evaluation Framework

A series of experiments assessed the platform's performance using Quality of Service (QoS) and Quality of Experience (QoE) metrics to identify optimization opportunities.

QoS Metrics

QoS provided a technical baseline for evaluating system performance:

- Response Delay (RD):

    - Processing Delay (PD): Time for the server to process and respond to user commands.
    - Playout Delay (OD): Time for the client to render and display frames.
    - Network Delay (ND): Round-trip time for data exchange between server and client.
- Network Load: Throughput, packet loss rate, frame rate, and traffic control.

Limitations: QoS focuses on technical aspects but overlooks user perceptions of performance.

QoE Metrics

QoE considers the subjective impact of network behavior on user satisfaction, providing a holistic performance view. It captures how network imperfections, such as latency or jitter, affect usability.

Experimental Setup and Key Findings

Initial Latency Testing

Two participants with HMDs connected to separate edge nodes participated in VR sessions. Key setup details included:

- Edge Nodes: Each ran signaling servers and LS components. Connected via 1Gbps Ethernet.
- HMDs: Connected via 5GHz Wi-Fi on 5G routers (VR HMDs in market do not provide 5G interface).
- Relay Server: Photon relay server deployed for coordinating session data.
- Data Collected: Metrics included message counts, byte rates, RTT, FPS, and transformation updates.

Results:

1. Framerate Dependency: Applications generated data based on frame rates, making high FPS (>65) critical for smooth collaborative experiences.
2. RTT Insights: RTT averaged ~70ms but increased due to 5G network congestion, highlighting the need for network optimization.
3. Message Optimization: Photon API effectively bundled queries, reducing message counts without losing data fidelity.

Enhanced Edge Node Testing

To improve system performance, three participants connected to GPU-enabled edge nodes through Geneva's 5G network. Key changes included:

- Dedicated edge nodes for each participant.
- Real-time VR movement tracking using the Medical Sample App.

Metrics Captured:

1. Render Time: Time per frame, with a target of ≤20ms.
2. Frames Per Second (FPS): Consistent values close to 60 FPS were achieved, a significant improvement from previous cycles (~30 FPS).
3. Network Stability: Byte rates stabilized after initial fluctuations, reflecting improved session reliability.
4. Battery Consumption: Average consumption per session was within acceptable ranges (4-14%), with higher usage correlated to lower initial battery levels.

Findings:

- Collaborative Mode Stability: Improved edge node applications ensured smoother VR interactions, even under varied network conditions.
- Latency and RTT: Average end-to-end latency stayed below 100ms, meeting the acceptable threshold for VR applications.

Key Metrics Evaluated Across Sessions

The table below summarizes critical performance parameters and their relevance:

Metric	**Description**
Render Time	Milliseconds per frame, critical for visual quality and responsiveness.
FPS	Frames per second, ensuring immersive and fluid VR experiences.
Round Trip Time (RTT)	Network latency from HMD to LS and back.
Packet Loss	Percentage of data packets lost during transmission.
Jitter	Variations in packet delivery timing, affecting real-time interactions.
Battery Consumption	Energy usage per session, crucial for usability during prolonged sessions.
Message and Byte Counts	Total and rate-based counts of messages and data exchanged during the session.

Latency and FPS Highlights

- **Latency:**
  - RTT values averaged ~70ms across sessions, with occasional spikes due to non-optimized 5G network parameters.
  - End-to-end latency remained below 100ms for most users, ensuring acceptable VR interaction quality.
- **FPS Improvements:**
  - Consistent FPS (~60) across tests marked a significant improvement from earlier cycles, ensuring smoother user experiences in collaborative modes.

Insights and Recommendations

The experiments highlighted the importance of robust networking for collaborative VR applications. Key takeaways include:

- High FPS is Crucial: Maintaining a consistent frame rate ensures smooth interactions and enhances user satisfaction.
- Optimized Network Parameters: Future improvements in 5G slicing and relay server deployment can mitigate congestion and improve latency.
- Efficient Resource Allocation: Dedicated edge nodes tailored to session requirements reduce latency and optimize system performance.

This task demonstrates a scalable, high-performance networking architecture capable of supporting immersive and collaborative VR applications, with future enhancements focused on latency reduction, resource optimization, and user satisfaction.

- **Deliverables**
  - D3.2: Networking scheme report (UNIGE, ORAMA, M15) - This report provided a description and evaluation results of the different deployment strategies as well as the final deployed networking scheme in detail.
- **Milestones**
  - MS3: Deployment of edge server
    - Goal 6: 5G enabled edge server (D3.2) - Completed

WP4 MAGES SDK physics-based cutting & tearing engine based on Geometric Algebra

Task 4.1: All-in-one Geometric Algebra engine (UNIGE, ORAMA)

Task Status: Completed

The MAGES platform incorporates an advanced Geometric Algebra (GA) engine to facilitate high-performance, real-time transformations and deformations critical to interactive medical training. This GA engine enables the system to execute complex, dynamic simulations while maintaining network efficiency and visual fidelity.

Key features of the GA engine include Efficient Network Data Transmission and GA-Driven Deformable Animations. Through compact GA-based representations, MAGES minimizes the volume of transformation data required for network transmission, reducing bandwidth demands by up to 33%. This efficient data handling allows MAGES to support high-fidelity, real-time interactions even on limited network infrastructures, enabling users to experience uninterrupted, accurate simulations in VR. The GA engine also supports advanced deformable animations for soft-body models, allowing for realistic cuts, tears, and physical deformations in the virtual environment. This feature is crucial for simulating tissue manipulation, as it responds dynamically to trainee interactions, creating an immersive, anatomically accurate surgical experience.

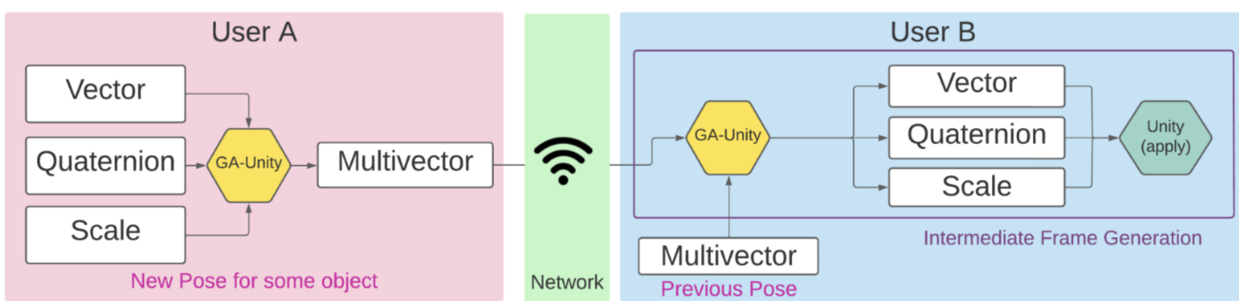

Figure 10 - An example of usage on how the data is compressed into multivectors and broadcasted over the network.

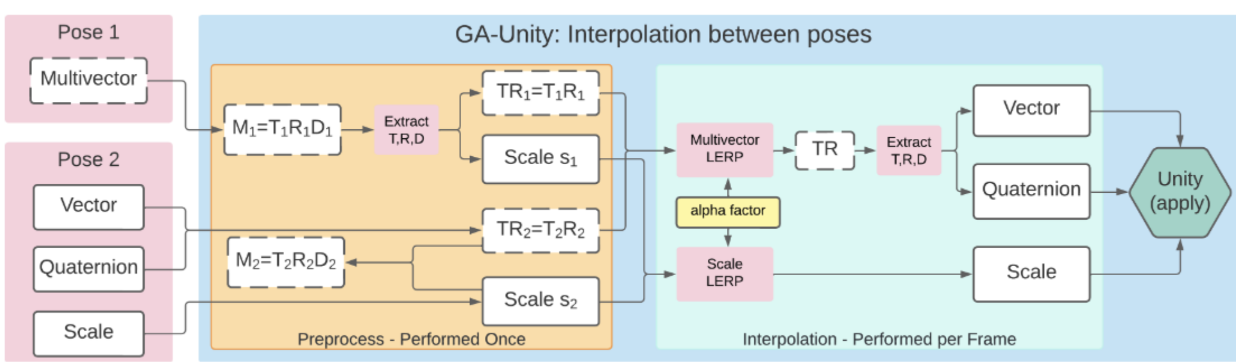

Figure 111 - An overview of the proposed pipeline for interpolating between two poses is provided. A pose can be input as a single multivector M, or as three multivectors T, R, D, or via its typical representation form, consisting of a vector, a quaternion, and a scale factor. Following a "preprocess" step to extract the T R and scale factor of each pose (if necessary),

corresponding vectors, quaternions, and scales can be obtained for each α (alpha factor), which are natively used by Unity. Dashed boxes indicate multivector forms.

Through this all-in-one GA engine, the MAGES platform enhances both the performance and realism of the IDS system, allowing trainees to experience responsive, immersive simulations of surgical tasks while benefiting from MAGES' robust and scalable architecture.

Task 4.2: Physics-based cutting and tearing engine (UNIGE, ORAMA)

Task Status: Completed

Within the MAGES platform, the physics-based cutting and tearing engine is crucial in simulating realistic, responsive surgical interactions in virtual environments. Figure 12 depicts the cutting process.

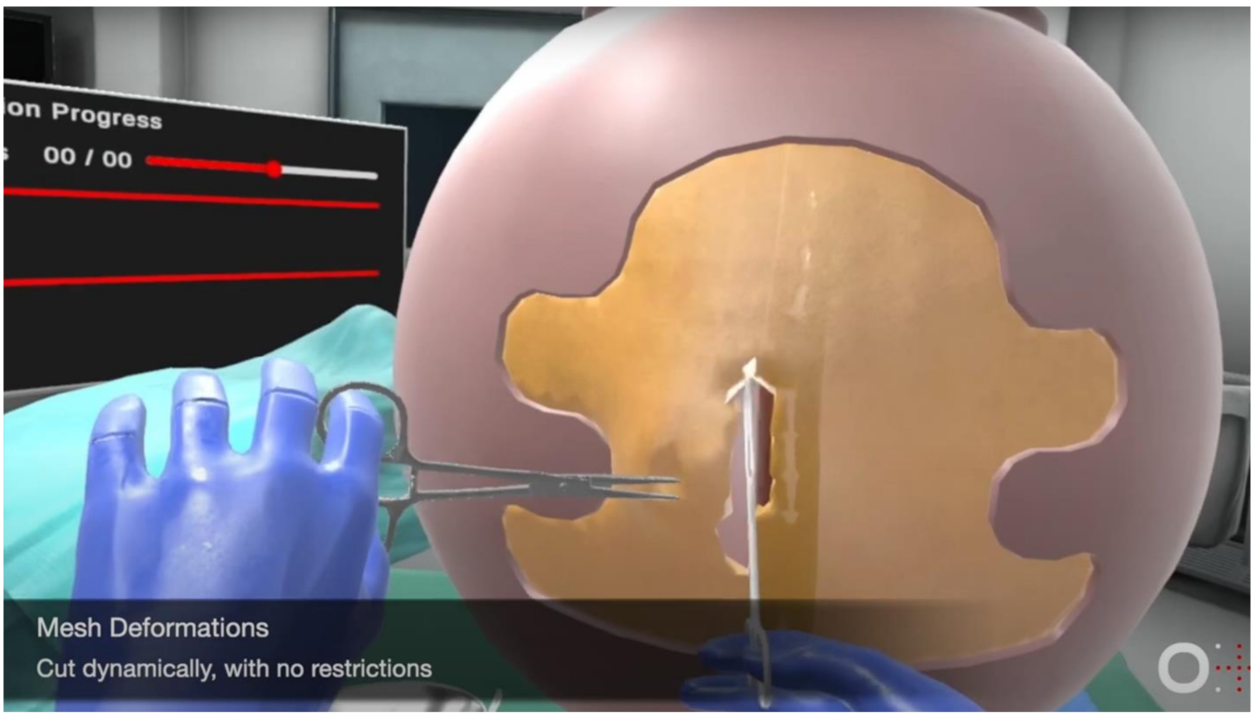

Figure 12 - Simulation of a dynamic cutting action within the MAGES platform, demonstrating unrestricted mesh deformations. The virtual tissue responds realistically to the instrument's movement, allowing precise incision that mirrors real-world surgical interactions.

We implemented our own physics system to simulate the soft properties of objects that were used in this project to simulate the human skin with all the different properties and physics materials. In this section we will present our algorithm and give an example of how a user can utilize our system.

Finally, the engine's Real-Time Performance Optimization ensures smooth rendering of complex interactions, achieving high frame rates even on VR devices with lower processing power. This optimization makes MAGES accessible across diverse hardware setups (contributing to our surgical education democratization efforts), providing a reliable VR experience that retains visual and mechanical fidelity. Through this physics-based engine, MAGES enhances the realism of the IDS system, allowing trainees to perform surgical procedures while receiving real-time feedback on technique, precision, and theory.

- **Deliverables**
    - D4.1 MAGES SDK enhanced release (ORAMA, M15) - Final release of the MAGES SDK component that implements the advanced Geometric Algebra engine, and cut/tear/drill surgical actions on deformable surfaces. You can download the latest version of MAGES SDK from here.
- **Milestones**
    - MS4: Components-microservices Implementation
        - Goal 7: Enhanced MAGES SDK component implemented and tested (D4.1) - Partially completed, final testing at HUG will be conduct after we collect trainee feedback and experience (awaiting for ethical committee approval, request done on October 1, 2024).

WP5 System Integration, Training module development and evaluation

This WP focuses on the unification, application, and assessment of the IDS system within the MAGES VR platform. We aimed to integrate the individual components developed in previous WPs into a cohesive training system and evaluate the effectiveness and usability of the system through real-world testing and user feedback. To achieve this goal, we divided WP5 into three main tasks: 1) Component Integration; 2) VR Training Use Case Module Development; and 3) User Evaluation.

Task 5.1: Component Integration (ORAMA, UNIGE, MIRALab SARL)

Task Status: Completed

This task is responsible for bringing together each component described to build the IDS system into a unified platform within the MAGES VR environment. This integration involved connecting the 3D animated surgeon (Task 2.1) with both the gesture recognition module (Task 2.2) and the feedback and recommendation module (Task 2.3), enabling real-time assessment and guidance. The 3D animated surgeon provides visual demonstrations of correct techniques based on data processed by the gesture recognition model, while the feedback module, powered by LLMs, delivers verbal explanations and clarifications to enhance learning retention.

The edge computing and 5G network support component ensures low-latency operation, allowing the IDS system to respond immediately to trainee actions within the VR environment (Tasks 3.1 and 3.2). This real-time responsiveness is crucial for effective training, as it enables immediate correction and learning reinforcement. Then, the physics-based cutting and tearing engine allows trainees to engage in realistic, physics-driven interactions with virtual tissues, such as making incisions or suturing, which adds a layer of practical skill application to the training module (Tasks 4.1 and 4.2).

The final integration of these components results in the IDS system deployed within the MAGES VR environment, providing a comprehensive, high-fidelity simulation for surgical training. This integrated system is designed to offer both immediate feedback on performance and a continuous learning environment, where trainees can refine their skills in a safe, immersive, and realistic setting.

- **Deliverables**
    - D5.1: Integrated IDS platform (ORAMA, UNIGE, MIRALab SARL, M19). You can download the latest version of MAGES SDK from here.
- **Milestones**

- MS5: Application integration
    - Goal 9: Final IDS application release (D5.1) - Completed

Task 5.2: VR training use case module development (ORAMA, HUG)

Task Status: Completed

For this work package, we utilized the softbody algorithms and MAGES SDK to create a VR training simulation for performing a craniotomy and sutures on the human skull. Our research pipeline was the following.

We collected visual material from a craniotomy operation from HUG to have it as reference since we wanted to simulate the human skull tissue as realistic as possible. The skin on the human skull has different tension that other parts of the body as it is more stretched due to the bone underneath.

Additionally, we acquired a 3D scanned model of a skull to implement our solution in VR. However, we had to heavily reduce the model polygons to make it affordable to run on a standalone VR headset. Initially, the model was close to 2 million faces, and we reduced it to 20K faces.

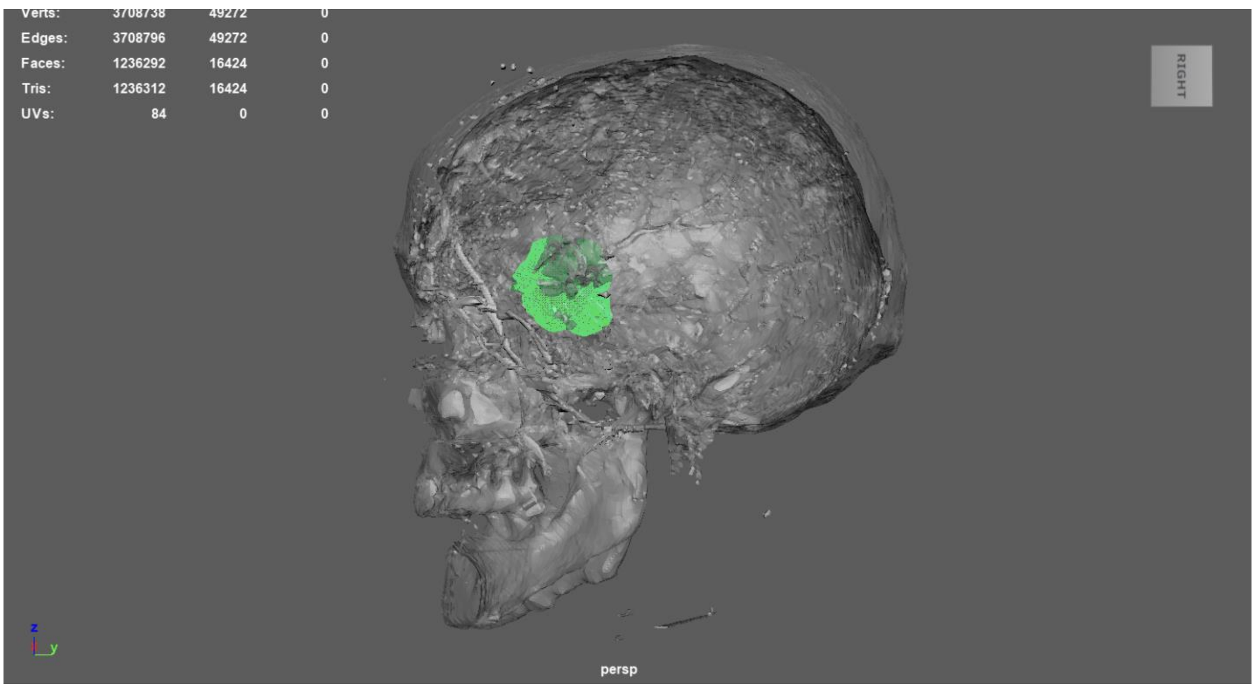

Figure 22 - The green highlighted object is the tumor that needs to be removed.

Then we imported the 3D model to the virtual environment, took the scalpel and performed the incision. We experimented with various softbody configurations till we had a satisfactory result.

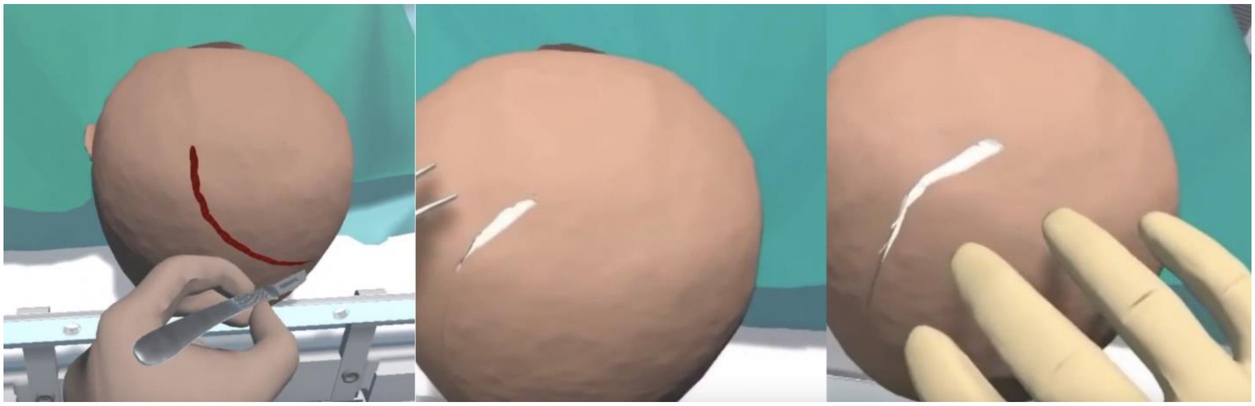

Figure 22 - The first image shows the incision on the head where blood was coming out of the skin. The second and third images follow a different variation and as we see the skin is stretched using pliers and the hands of the doctor. The skin behaves as it should in a real scenario with the proper tension and flexibility.

We applied the same methodology to simulate a different scenario in hemihepatectomy where the user should remove the damaged part of a liver. In this case we configured the liver to be more stiff than the skin since it behaves in a different way.

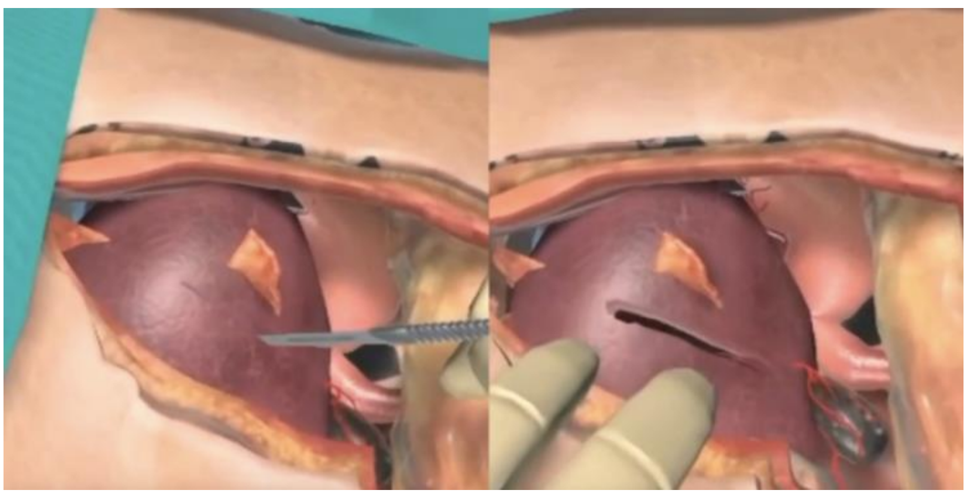

Figure 23 - In the first image we see the surgeon cutting the liver with the scalpel. The second image the surgeon interacts with the deformable liver by touching it and we can see that the organ behaves accordingly.

Following the interactions with human organs and skin the next task we had to implement was to perform sutures. We made this possible through our mesh deformation system.

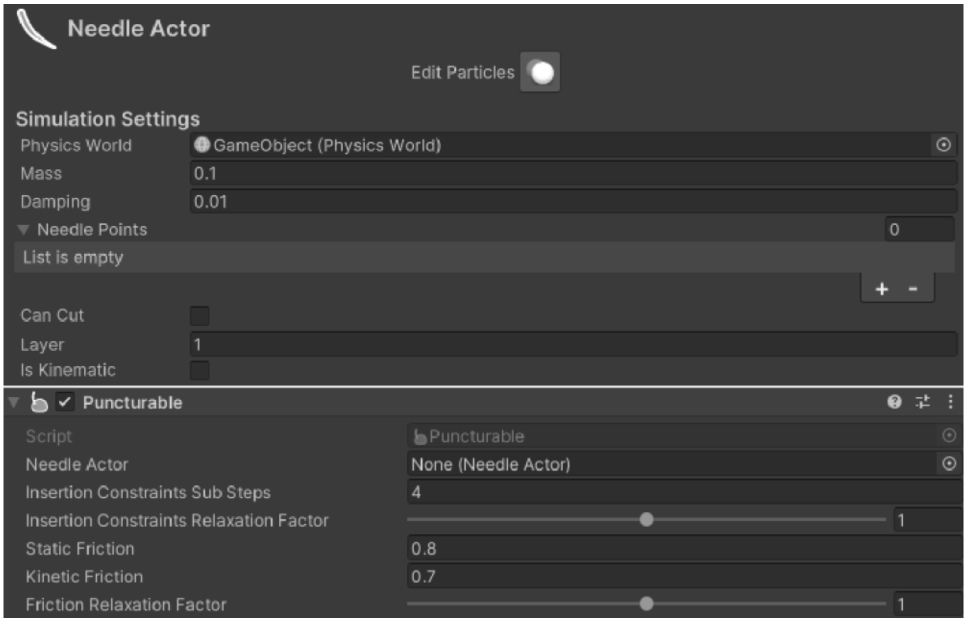

Figure 24 – The needle actor and the puncturable components are both needed to perform the suturing. The first one is attached to the 3D model of the needle whereas the second one is attached to the surface that will be sutured (e.g. skin).

To test our system, we utilized a different use case. We created a robotic simulator that performs sutures over a deformable area similar to a fake skin for suturing training. The surgical robot we created has the same principles as the robotic system of Davinci where the user can grab the masters that control the robotic hands and perform the suturing.

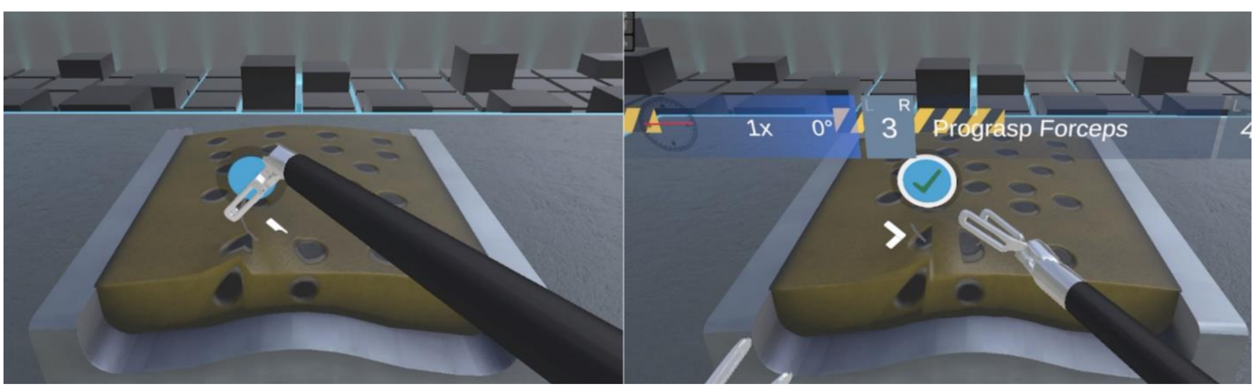

Figure 25 – Those figures show the different stages of the suturing where the surgical arms hold the needle, and they push it through the fake skin. The physical deformation that is applied to the fake skin is visible and caused by the forces of the needle to puncture its surface.

- **Deliverables**

- D5.2: Incision-Suturing training module (ORAMA, HUG, M21) A video with the result application in comparison with the real video footage can be seen here.

- **Milestones**
  - MS6: User Evaluation
    - Goal 10: Training module development (D5.2) - Completed

Task 5.3: User evaluation (HUG, ORAMA)

Task Status: Partially Completed

To validate our results regarding the deformable meshes, our cut and tear algorithms and the final VR application that we built for the craniotomy, we conducted a preliminary qualitative user evaluation on our premises.

In this section we will present our results about those 3 testing sessions.

Deformable meshes

The evaluation involved 10 developers with varying levels of experience-four junior, four intermediate, and two senior participants. They were asked to integrate deformable functionality into three 3D models: a rubber ball, a human liver, and a surgical cushion, using the MAGES SDK. The developers followed our provided documentation, which included written guides and code examples. Eight out of ten participants completed the task successfully, with a median completion time of 30 minutes, ranging from 15 to 40 minutes.

Several challenges were encountered during the task. Four developers noted gaps in the documentation, particularly the absence of troubleshooting tips for common errors. Shader integration, for instance, posed significant difficulties for two participants due to missing guidance on compatible render pipelines. Additionally, two developers experienced crashes when working with larger meshes, which interrupted their progress and required them to restart their workflows.

Feedback from the developers highlighted both strengths and weaknesses. While the documentation was generally appreciated for its structure, participants recommended including visual examples and video tutorials to clarify core concepts. They also suggested enhancing error messages to provide more detailed explanations. Despite the issues, the developers rated the tool’s ease of use at an average of 3.8 out of 5.

Five VR end-users participated in testing the deformable meshes using an Oculus Quest 3 headset. Their tasks included pressing, stretching, and manipulating the 3D models to evaluate the system’s responsiveness and realism. The success rate among this group was mixed, with three users completing the tasks without significant assistance, while two struggled with understanding the interactions.

The testing revealed notable issues with the system’s feedback and responsiveness. Three users reported that the deformation behavior was inconsistent with their expectations. For instance, one user remarked that the liver model felt more like a rubber object than soft tissue. Four participants found the interaction controls unintuitive, particularly those unfamiliar with VR applications, who struggled to understand how to grab or stretch the meshes. The absence of tooltips or an in-app tutorial exacerbated these challenges. Additionally, two users experienced noticeable performance drops during rapid manipulations of larger meshes, which affected their overall impression of the system.

In terms of realism, users gave the visual quality of the meshes a favorable rating, but their behavior under interaction received less enthusiasm. Participants suggested enhancing the realism of the deformation through improved haptic feedback and optimizing performance to ensure smoother interactions. The average realism rating for appearance was 4.1 out of 5, while behavior scored 3.2 out of 5.

Cut and tear algorithms

The evaluation of the cut and tear algorithms involved the same group of ten developers who participated in the deformable mesh testing. They were tasked with extending the functionality of the same three 3D models by making them capable of being cut and torn using tools provided by the MAGES SDK. Developers were given access to online documentation and tutorials, similar to the deformable mesh evaluation.

Seven developers successfully completed the task, with an average completion time of 30 minutes. Two developers struggled with configuring the cutting tools to recognize the mesh boundaries properly. They reported that the available parameters were unclear, leading to trial-and-error adjustments that prolonged their workflow.

Four participants commented on the overall complexity of the toolset, stating that the abundance of configuration options made it difficult to identify the essential parameters. This sentiment was echoed by others, who suggested that pre-configured templates or simplified options could make the tool more accessible, particularly for those less familiar with the underlying mechanics.

Despite these challenges, several developers praised the overall power of the tools once configured correctly. One senior developer noted that the algorithms performed impressively during testing, but they required a steep initial learning curve to fully utilize.

The five VR end-users were tasked with evaluating the cuttable and tearable meshes created by the developers. Using an Oculus Quest 3 headset, they performed actions such as making incisions on the liver model and tearing the surgical cushion. The results varied widely depending on the user's familiarity with VR and the system's responsiveness.

Two participants struggled with understanding the depth required to make accurate incisions. This lack of depth perception was a recurring issue, leading to imprecise cuts that didn't align with their intended paths. Additionally, the absence of haptic feedback made the cutting and tearing feel less immersive, detracting from the overall realism of the experience.

The remaining three users, who were more experienced with VR applications, were able to perform the tasks successfully but still identified areas for improvement. One participant suggested that the tearing behavior felt unnatural, noting that the material didn't respond realistically to their applied forces. Another user observed performance drops during the tearing action, especially when applied to more complex meshes.

Craniotomy VR simulation

Performance among the participants varied significantly. Five users achieved a score of 70%, while four scored 60%, and six scored 50%. Several factors contributed to these results, with the most common issue being the lack of depth feedback during the incision process. Three participants noted that it was difficult to judge how deep they were cutting, leading to errors that impacted their scores.

Four participants reported challenges with the controls, particularly with the scalpel's movement and positioning. The lack of an onboarding tutorial within the VR application left some users struggling to understand the interaction mechanics.

Another issue identified was the system's lack of tactile or visual feedback during the incision process. For example, two users mentioned that the simulation did not provide adequate cues, such as resistance or visual changes, when cutting through layers of the virtual scalp or skull.

Despite these challenges, most participants agreed that the simulation had significant potential as a training tool. The immersive environment and the detailed anatomical models were praised for providing a realistic representation of a craniotomy procedure.

While the deformable mesh and cut-and-tear features demonstrated significant potential, the evaluation revealed areas for improvement, particularly in user interaction and system feedback. The craniotomy simulation, though functional, highlighted key areas such as depth visualization and control intuitiveness. Future iterations will prioritize addressing these issues to ensure a seamless experience for both developers and trainees.

- **Deliverables**
    - D5.3: User Evaluation report (HUG, ORAMA, M24) (Partially Completed)
- **Milestones**
    - MS6: User Evaluation
        - Goal 11: User evaluation report (D5.3) - Partially Completed, the user evaluation protocol is defined. However, we have not finalized the evaluation in the use case at HUG.

Summary

This section provides an overview of the primary work packages (WPs), tasks, deliverables, milestones, and completion status for the IDS project, highlighting the collaborative contributions of MIRALab, UNIGE, HUG, and ORamaVR. Each WP addresses distinct components of the IDS system, from defining requirements and architecture to designing and implementing real-time gesture recognition, feedback systems, edge computing, and VR-supported modules. Table 3 consolidates the information provided in response to Question 1.1.

Work Package	Task	Deliverables	Milestones	Status
WP1	Task 1.1	D1.1	MS1	Completed
	Task 1.2	D1.2	X	Completed
WP2	Task 2.1	D2.1	X	Completed
	Task 2.2	D2.2, D2.3	X	Partially
	Task 2.3	D2.4	MS4	Partially
WP3	Task 3.1	D3.1	MS2	Completed
	Task 3.2	D3.2	MS3	Completed
WP4	Task 4.1	X	X	Completed
	Task 4.2	D4.1	MS4	Partially
WP5	Task 5.1	D5.1	MS5	Completed
	Task 5.2	D5.2	MS6	Partially
	Task 5.3	D5.3	MS6	Partially

Table 1 - Summary of Work Packages, Tasks, Deliverables, Milestones, and Status.

In summary, the IDS project has significantly contributed to advancing surgical training in the VR environment. We have defined comprehensive use cases and requirements for effective gesture training, developed a cognitive digital surgeon that combines real-time gesture recognition and feedback through deep learning algorithms and Large Language Models (LLMs), and implemented a 3D animation engine to simulate surgical actions accurately. Additionally, the integration of edge computing ensures responsive performance in VR environments, and the MAGES SDK offers advanced physics-based interactions for realistic surgical training. While final validation with medical students at HUG awaits authorization from UNIGE's ethical committee, this project has already established a robust foundation in intelligent surgical training, with each completed component moving closer to the goal of accessible, high-quality medical education technology.

Future Work

Additional Surgical Gestures

Currently, our system focuses on two surgical gestures: incision and suturing. However, future versions of the system could expand to cover a broader range of surgical tasks, including more complex gestures. For instance, by training the gesture recognition module on additional datasets, the system could be adapted to cover areas such as laparoscopic procedures, orthopedical surgeries, and even microsurgery.

Different Aspects of a Surgery

In this project, we focused on recognizing and explaining the gestures performed by trainees. However, future versions of the system could extend this capability to assess the outcome of these gestures. For example, it would be valuable to evaluate not only the sequence of steps for suturing but also the quality of the suture itself. By developing algorithms that assess the results of surgical tasks, the system could provide even deeper and more comprehensive feedback, enabling trainees to understand how their techniques translate into successful or unsuccessful surgical outcomes.

Enhanced Developer and User Experience

We aim to further improve the MAGES platform by enhancing both developer and user experience. Future updates will focus on facilitating the usability of MAGES to the point where even developers with minimal experience can efficiently create and deploy VR training simulations. With intuitive workflows and simplified access to cloud analytics, the goal is to allow developers to set up comprehensive training scenarios in a single day. This enhancement would significantly reduce the learning curve and make VR training simulations more accessible across educational institutions.

Machine Learning Training

While the current system employs supervised learning, future work could explore alternative machine learning approaches, such as weakly-supervised learning or even neuro-symbolic paradigms. These methods could reduce the reliance on large, labeled datasets, which can be time-consuming and expensive to collect. By employing different ML approaches that require less extensive data labeling, our system could achieve improved performance with fewer resources, making it even more scalable and adaptable to new surgical tasks or training environments.

New Domains

While the IDS was developed with surgical training in mind, the underlying technologies can be applied to other domains that require precise movements and skill assessment. In this context, two particular applications seem relevant to tackling societal challenges. First, our system could be adapted to monitor patients' recovery and motor skill improvements in a physical therapy context, providing feedback to both patients and therapists. Second, gesture recognition could be used to analyze and improve athletic performance in sports requiring precise movements, such as gymnastics or martial arts.

Interoperability with Universal Scene Description (USD) Format

To facilitate collaboration across different 3D platforms and new domains, future versions of MAGES plan to support the Universal Scene Description (USD) format. USD is widely adopted as a standard for interoperable 3D assets and integrating it within MAGES would expand collaborative possibilities by allowing assets to be easily shared and used across other virtual environments. This development would enhance the cross-platform versatility of MAGES, making it easier for educational and training organizations to incorporate VR into existing systems.

Question 1.2 - Have scientific publications, inventions or patent applications been made? If yes, please give details regarding the status.

The following scientific publications have resulted from the research and development activities within the Intelligent Digital Surgeon (IDS) project. Moreover, we include the next planned scientific publications.

- *TransSG: A Spatial-Temporal Transformer for Surgical Gesture Recognition* **(Published)**
  https://intelligentdigitalsurgeon.miralab.com/wp content/uploads/2024/12/CGI2024_127.pdf

- *MAGES 4.0: Accelerating the World's Transition to VR Training and Democratizing the Authoring of the Medical Metaverse* **(Published)**[3]
  https://ieeexplore.ieee.org/document/10038619

- GA-Unity: A Production-Ready Unity Package for Seamless Integration of Geometric Algebra in Networked Collaborative Applications **(Published)**[4]
  https://www.kamarianakis.eu/publication/2024-gaunity-engage/

- Progressive tearing and cutting of soft-bodies in high-performance virtual reality **(Published**[5]
  https://diglib.eg.org/items/01417f77-73bd-428b-8779-2ea664a23d2f

- VR Isle Academy: A VR Digital Twin Approach for Robotic Surgical Skill Development **(Published)**[6]
  https://arxiv.org/abs/2406.00002

---

[3] Zikas, Paul, et al. "MAGES 4.0: Accelerating the world's transition to VR training and democratizing the authoring of the medical metaverse." IEEE Computer Graphics and Applications 43.2 (2023): 43-56.

[4] Kamarianakis, M., Lydatakis, N., & Papagiannakis, G. (2024). GA-Unity: A Production-Ready Unity Package for Seamless Integration of Geometric Algebra in Networked Collaborative Applications. *arXiv preprint arXiv:2406.11560*.

[5] Kamarianakis, M., Protopsaltis, A., Angelis, D., Tamiolakis, M. & Papagiannakis, G. (2022). Progressive tearing and cutting of soft-bodies in high-performance virtual reality. arXiv: 2209.08531, [cs] (2019) (available at https://arxiv.org/abs/2209.08531) to also appear in ICAT-EGVE 2022 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments. The Eurographics Association."

[6] Filippidis, A., Marmaras, N., Maravgakis, M., Plexousaki, A., Kamarianakis, M., & Papagiannakis, G. (2024). VR Isle Academy: A VR Digital Twin Approach for Robotic Surgical Skill Development. arXiv preprint arXiv:2406.00002.

- *"KD-MVNet: Knowledge Distillation Multi-View Network for Single-View Surgical Gesture Recognition"* **(Planned)**

  This publication will present the implementation of the core framework, including multi-view training pipeline, spatial-semantic feature extraction, and initial knowledge distillation setup. The described experiments will include knowledge distillation training, single-view inference testing, and various ablation studies to validate our methods' effectiveness.

- *Virtualize Your Training: Integrating VR Systems into Surgical Education—A Case Study in the Hospital Domain* **(Planned)**

  This publication will describe the evaluation process of the IDS system as a training tool within a hospital-based teaching environment. The described experiments will be conducted in a controlled hospital setting, inviting surgical trainees at different levels to perform a designated surgical task using the IDS system.

- *Benchmarking Gesture Recognition in Surgical Training: Insights from the CGI 2025 Competition* **(Planned)**

  This paper will present and analyze the outcomes of the CGI 2025 competition, focusing on evaluating diverse gesture recognition approaches in surgical training. The competition will invite researchers to develop and apply different algorithms to an open surgical gesture dataset, comprising expert and student demonstrations captured as part of the Intelligent Digital Surgeon (IDS) project.

- *Adapting Large Language Models with Retrieval-Augmented Generation for Multi-Modal Surgical Training: A Case Study in Medical Knowledge Integration* **(Planned)**

  This paper will propose a novel approach integrating Large Language Models (LLMs) with a Retrieval-Augmented Generation (RAG) framework to deliver precise, contextually relevant feedback for surgical training in a VR environment. The study will detail the adaptation of LLMs to a medical knowledge base provided by HUG, containing structured procedural guidelines, terminologies, and best practices in surgical training.

Question 1.3 - Summarize the progress of the activities in relation to the implementation plan and assess the commercial, economic and/or social benefit.

Status of the implementation plan:

The IDS project is currently on track, with the core technical components developed and the majority of milestones achieved within the planned timeline. The base components include the animated 3D intelligent digital surgeon, the real-time gesture recognition model, the LLM-based recommendation system, the edge computing and 5G network support, and the geometric algebra and physics-based engines, all integrated into the MAGES platform.

What implementation scenario is being considered?

The primary implementation scenario envisions IDS as a modular VR-based training tool that medical institutions can integrate into their existing curricula to support surgical training. Targeted for resource-rich, limited environments, and diverse regions, the IDS platform is designed to be deployed with portable VR equipment, thereby broadening access to high-quality training. Additionally, there is potential to adapt the platform for other professional training domains, such as physical therapy and sports, due to the underlying gesture recognition and feedback capabilities.

What are the main milestones and decision-making points for implementation?

We have defined different decision-making points that align with quantitative assessments to track progress, such as user performance improvements, accuracy of gesture recognition, and user satisfaction scores. Then, we defined the following milestones for implementation:

- **Pilot testing and validation at HUG:** The feedback from medical trainees will inform final adjustments to improve usability and effectiveness.
- **Intellectual Property Rights (IPR) and Commercial Readiness:** Ensuring intellectual property protection and establishing licensing agreements.
- **Market Readiness for Expansion:** After the system proves successful in surgical training, decisions will be made on expanding to other areas, such as athletic training and physical therapy.

How is the necessary know-how transfer ensured? How were the implementation partners been able to strengthen and expand its knowledge, technology, and innovation base through cooperation with research partners?

In ensuring the necessary know-how transfer, the IDS project established a structured collaboration framework among all partners, supporting continuous knowledge sharing and skill development for successful implementation. Regular meetings were essential, facilitating open communication and timely updates across MIRALab, UNIGE, HUG, and ORamaVR.

The collaboration between MIRALab and UNIGE was fundamental to this process, with both institutions actively engaging in 2 PhD student supervision and structured mentoring sessions. These sessions enabled a thorough transfer of technical expertise in areas like AI model development, gesture analysis, LLMs, and VR integration.

Additionally, HUG's medical staff provided essential guidance on medical practices and training needs, ensuring the IDS development stayed aligned with real-world medical training demands. Their input shaped the design of data capture, gesture recognition, feedback mechanisms, and interface elements to meet the practical requirements of surgical training for medical students.

ORamaVR also maintained a close working relationship with HUG, reflecting a history of continuous collaboration on various projects and benefiting from recurrent meetings to align on project needs. Moreover, ORamaVR worked closely with UNIGE and MIRALab, where professionals exchanged insights into leveraging VR environments and transferring techniques from real-world settings into the MAGES platform. This exchange of knowledge was invaluable for adapting complex surgical interactions into VR, enabling the integration of real-world requirements into the digital training experience provided by IDS.

What quantitative economic results can be expected from the implementation? The economic benefits are expected to materialize primarily through:

- **Cost savings:** Educational institutions and hospitals can reduce their training costs by adopting a platform that does not need a physical training setup with costly equipment.
- **Time savings:** Instructors can save time by not having to specifically look at every single step of a training session, focusing only when the IDS finds a mistake and evaluating the problematic steps. The trainees can save time by having their training sessions anywhere and anytime, without the need to rely on the availability of professionals.
- **Revenue generation:** Our collaboration can have access to funding by licensing the IDS technology to educational institutions, medical institutions, and companies that

build products for gesture recognition and interactive feedback. There are different domains in which the IDS technology could be deployed and generate revenue for the partners.

Have new risks been identified which could affect the future commercialization activities?
Through our development process, we have identified new potential risks:

- **Institution adoption barriers:** Some institutions may face challenges in integrating VR solutions due to technological or budget constraints.
- **Data privacy concerns:** Given the personalized feedback system, ensuring compliance with data privacy regulations will be essential, particularly in medical training contexts. Moreover, since we train ML models, it is necessary to adapt constantly to the different regulatory bodies, like the European Union.
- **Technological dependencies:** As the system relies on edge computing and 5G support, any limitations in network infrastructure could affect performance in less developed regions.

To mitigate these risks, IDS will focus on modular deployments, alternative compliance pathways for data privacy, and further performance optimization for low-bandwidth environments.

Refer to the measurable and quantified objectives of each milestone and provide data supporting your achievements

- Gesture Recognition Accuracy (Milestone: Algorithm Development and Testing): Gesture recognition accuracy achieved 83.6% on initial tests.

Question 1.4 - What is the implementation partners’ strategy for (further) scaling and growth?

Are follow-up activities from the Innosuisse project planned, and if so, to what extent and which persons and/or partner/s will be involved?

Yes, follow-up activities from the Innosuisse project are planned. Building on the success of the Intelligent Digital Surgeon (IDS) project, we aim to expand the innovations and systems developed within the consortium.

We intend to complete the collection of trainee gesture datasets and enhance the AI models for better recognition accuracy and feedback mechanisms. The system will be adapted to include more complex surgical gestures, addressing more medical sectors and disciplines.

Each partner has the following action items planned:

Hôpitaux Universitaires de Genève (HUG): Continued collaboration for clinical validation and feedback on medical training applications.

University of Geneva (UNIGE): Development and refinement of gesture datasets, AI algorithms, and VR system design.

MIRALab SARL: Focus on creating realistic 3D models and animations for additional surgical gestures.

ORamaVR: Optimization of the deformable softbody system to support multiple actors in parallel as well as commercialization efforts of the system.

The project has also inspired a follow-up European project, INDUX-R, currently ongoing with multiple European partners. This collaboration focuses on implementing new ideas and technologies derived from the IDS project to further innovate VR-based medical training.

How do you see the economic market potential as well as the value creation potential of both the project and the innovation in the long run?

The IDS project addresses critical gaps in surgical education by democratizing access to high-quality training, even in resource-limited settings. By reducing dependency on costly equipment and enabling training through portable VR devices, IDS offers a scalable solution for medical institutions worldwide. Its modular approach ensures adaptability to various training needs and domains, such as sports or physical therapy, further broadening its market applicability.

Key economic benefits include:

- **Cost Reduction**: Institutions can reduce the cost of maintaining physical training setups and expensive equipment by using IDS.
- **Revenue Streams**: Licensing the IDS technology provides opportunities to partner with educational and medical institutions globally.
- **Broad Market Appeal**: The IDS technology is adaptable for various applications beyond medical training, opening doors to other industries requiring precise skill assessments, such as sports or rehabilitation.

IDS lays a foundation for continuous improvement and innovation in VR-based medical training. By leveraging cutting-edge technologies like gesture recognition, machine learning, and real-time deformations, the system aligns with global efforts to standardize and elevate medical education. Its ability to improve skill acquisition, reduce training costs, and minimize medical errors contributes to:

- **Global Health Equity**: Offering affordable and accessible training solutions to underserved regions enhances healthcare quality globally.
- **Surgical Precision and Safety**: Better-trained professionals result in reduced surgical errors, benefiting patients and healthcare systems.
- **Sustainability**: Virtual training minimizes waste and reliance on physical materials, contributing to ecological goals.

We are confident that IDS will continue to evolve and expand its impact across multiple sectors, reinforcing the mission of each partner from the consortium to advance skill-intensive training through immersive technologies.

Question 1.5 - Describe the financial status of the project

The budget for the UNIGE partner is as following: The actual salary costs were of 505920.90 CHF (budget 506880 CHF), actual material costs 35561.00 CHF instead of the budgeted 30048 CHF, overhead costs were of 75888.13 CHF instead of 76032 CHF which gives a total actual cost of 617370.03 CHF instead of 612960 CHF. Therefore, the actual total cost of UNIGE amounts to 100.71% of the initially previewed amount. However, these numbers are still on a preliminary basis and might have slight changes on the consolidated report. For the HUG partner the salary costs were 67737.60 CHF with an overhead of 10160.64 CHF and a total of 77898.24 CHF which corresponds to 100 % of the original budget. MIRALab partner had a total cost of 142662.10 CHF while OramaVR reports that the full personnel effort of 3,439 hours was provided as committed. This was achieved using personnel with lower hourly rates than the initially estimated average of 58 CHF/hour, resulting in a total personnel cost of 169,070.35 CHF (70% of the originally estimated budget).

Question 1.6 - How do you assess the impact of the project in terms of ecological sustainability?

To assess the ecological and social impact of the IDS project, we can explore several dimensions that highlight how this project contributes to sustainable development and social equity. Next, we present aspects beyond democratizing technology and surgical training access.

- **Ecological Sustainability**
  - **Sustainable Resource Use:** Traditional surgical training setups, at every single educational institution, requires generating biological and synthetic waste, whereas VR-based training avoids this issue. By using a VR-based approach, the IDS project minimizes the need for physical materials traditionally required in surgical training, such as animal models (e.g., chicken legs, sedated mice), simulators, and consumables like suturing materials. Reducing dependency on these resources aligns with sustainable consumption practices, helping conserve medical supplies and reduce waste generation.
  - **Climate Change Mitigation:** The VR training platform, designed for portable VR devices with lower energy requirements, contributes to climate change mitigation efforts. Training in a virtual environment lowers the need for travel to centralized training facilities, which can reduce greenhouse gas emissions, especially for students in remote or underserved areas. Additionally, the implementation of edge computing allows for efficient data processing, lowering energy consumption in cloud infrastructures.
- **Social Sustainability**
  - **Democratizing Surgical Training:** IDS facilitates broader access to high-quality surgical training for students in low-resource settings and remote areas, who otherwise might not have access to advanced training or expert feedback. This democratization of technology reduces educational inequality by leveling the field between students in well-equipped hospitals and those in low-resource regions.
  - **Social Cohesion and Participation:** By integrating advanced, interactive VR technologies, the IDS project fosters a learning environment that encourages collaborative training and continuous peer feedback, even across geographically dispersed locations. This collaboration strengthens social

cohesion among medical professionals and students and promotes a shared culture of continuous learning and support.

  - **Reduction of Social and Gender Inequality:** Access to VR-based surgical training can help reduce barriers often experienced by disadvantaged social groups, including those in low-income areas. It provides equal access to training for men and women, ensuring gender equality in opportunities to acquire high-level skills necessary for surgical careers.
  - **Improvement in Quality of Education:** VR-based training offers immersive, repeatable, and error-tolerant learning experiences, contributing to higher-quality, practical surgical education. This training can improve the skill acquisition process, making it possible for medical students to practice until they reach proficiency without limitations on materials or time, which is often a constraint in traditional setups (i.e., professional surgeons are not always available to provide feedback and recommendations).
  - **Reduction of Social Costs:** With enhanced training accessibility, there is a potential long-term reduction in healthcare costs as a better-trained workforce translates into improved surgical outcomes, fewer medical errors, and shorter patient recovery times. Reducing medical errors also reduces associated social costs, including patient care costs, rehabilitation, and lost productivity due to surgical complications.
  - **Development Assistance and Cooperation:** IDS has the potential to extend its impact through partnerships with healthcare institutions in low- and middle-income countries. By fostering cooperation and supporting the training of healthcare professionals worldwide, IDS can promote sustainable development and build healthcare capacity in regions facing healthcare shortages.

In addition to addressing key challenges today with our proposed solution, we are also planning future steps to expand our impact in these areas:

- **Incorporating Low-Energy Hardware:** We plan to enhance the sustainability of the IDS project by prioritizing VR equipment with minimal energy requirements. Additionally, optimizing our algorithms to perform with lower computational power while preserving accuracy will make the solution more efficient and environmentally friendly.

- **Integrating Inclusive Training Data:** To broaden accessibility, we aim to incorporate training data that reflects diverse user needs. By adapting our gesture recognition and feedback systems to accommodate varying physical capabilities, we can make the IDS platform more inclusive and effective for a wider audience.

Question 1.7 - How successful was the project overall?

The project is highly successful. First of all, it has allowed a collaboration between HUG, UNIGE, MIRALab and ORAMAVR that was quite successful, and we have been able to finalize most of our deliverables. We are waiting for the ethical committee answer to finalize the study at HUG.

The surgeons we have worked with (in brain surgery) are very satisfied with the collaboration of UNIGE and MIRALab for translating their needs in new datasets and into the surgical surgeon. The implementation of all the methods into the ORAMAVR platform is very successful.

This project has opened new avenues of research. A EU project INDUX-R, a follow up on many new IDS ideas has been accepted that is now ongoing with European partners.