User interaction design in next-generation immersive systems

Aliaksandr Halavaty and Valeria Bahai
and Danila Bernat
Belarusian State University
Minsk, Belarus
Email: alex.halavatyi@gmail.com, lerika1307@mail.ru,
danila.bernat@gmail.com

Katsiaryna Halavataya
Email: kat.golovataya@gmail.com

***Abstract*—This article explores the issues of designing intelligent systems using augmented reality (AR) and virtual reality (VR) technologies. These technologies are based on the use of three-dimensional representation of the surrounding real or pre-designed virtual scene, the capabilities of these systems for determining the orientation of user and objects in space, intelligent analysis algorithms and machine learning. The complexity of developing individual algorithms, the need to take into account the specifics of the subject area for each applied system, interactive display based on threedimensional and other types of representations necessitates the development of new approaches and design principles. The article focuses on the capabilities of OSTIS Technology for use in these tasks, and also provides examples of building a system for technical description of equipment and a guided quest in augmented reality.**

***Keywords*—virtual reality, augmented reality, 3D representation, knowledge base**

I. INDRODUCTION

When designing virtual reality (VR) and augmented reality (AR) systems, it is necessary not only to create a high-quality image and show it to the user at the right time – it is necessary to design the user's environment, other characters, interaction with other characters and objects, and use informational environment and metadata to work with specific objects. The user must feel like a part of the space (in the case of virtual reality) or perceive the content in relation to the physical environment (in the case of augmented reality). At the same time, it is necessary to properly implement and design various optical and auditory effects, correctly track user movements with sufficient precision, implement ways to interact with various objects and subjects of the surrounding space. In addition, all visualisations, reactions, interactions and updates to the designed space and objects must be carried out in real time in order not to disrupt user's immersion and their psychoemotional state while using the system. This research area is related, among other things, to the field of immersive systems design. The paper [1] describes four different components of user interaction immersion: sensory-motor immersion, cognitive immersion, emotional immersion and spatial immersion.

Another recently popular concept is mixed reality (abbreviated XR). Mixed reality systems approach constructing immersive interactions by separating surrounding systems, subjects, objects and informational space into two different types of so-called world layers — physical world layer, containing objects that are naturally present and can be perceived by the user as separate entities, and informational world layer, that considers various types of information and data that can be used to design and enhance user's interaction with the system. Mixed reality strives to build a combination of objects from these two different world layers. At the moment, the concepts of virtual reality and augmented reality are most often used in applied systems [2].

Virtual Reality is a high-level user interface that includes real-time simulation and interaction through multiple sensory channels. In VR, the scene is constructed artificially and then perceived by the user by effectively overtaking their sensory channels (primarily visual and auditory) and supplying pre-designed data to these channels instead of the real world data.

The main differences between virtual reality and other interfaces that use the visual presentation of information to the user are:

- 3D stereo vision,
- user's viewpoint control – the system changes the visible display point of view in the constructed space by using user's viewpoint and view orientation changes in the real world as input,
- possibility of interaction with the virtual environment in real time.

Augmented Reality is the result of introducing any additional constructed types of visual representation of data into the regular visual information feed, in order to provide additional information about the environment and change the user's perception of this environment.

Promising applied areas in which the number of virtual and augmented reality systems is expected to grow, in addition to gaming and entertainment industry, include healthcare and medical devices, education, staff development and training, manufacturing, automotive industry, marketing and advertising, logistics and transport, retail, scientific visualization.

The main advantages of technology include:

- Visualization and interaction with objects and concepts that are hard or impossible to implement in the real world.
- Visualization of 3D concepts, modeling, viewing objects and scenes from different perspectives using a more natural way of observing them compared to traditional display devices.
- Modelling and research of hazardous and potentially dangerous environments, objects and situations.
- Promoting innovative styles and methods of study and learning.

II. Analysis of existing approaches

Despite the prospects and advantages of using these technologies, there are a number of limiting factors that significantly hinder all developments in the area and applying them to more real-world problems:

- High complexity of application systems development for VR, AR and XR. When designing a virtual reality application, it is necessary to consider the specifics of the target hardware optical system, user movement tracking sensors, and APIs for retrieving this information from the device, which can be fundamentally different for different manufacturers. When constructing a scene observed by the user, it is necessary to generate a high-quality image, use directional acoustic models, etc., which requires significant hardware processing capacity on the enduser device, or stable and very low-latency remote compute server link. Augmented reality systems are additionally based on a technical vision hardware and algorithms, and building such models also presents multiple challenges. All these factors mean that it is practically impossible to design such systems by individual developers, and developing such systems requires sufficient expertise.
- Insufficient amount of existing content. The lack of content, in addition to the complexity of developing individual systems, is also due to the fact that existing virtual reality systems are based on different hardware and application platforms, often tied to the physical characteristics of the processing system and end-user equipment, which makes them software or hardware incompatible.
- Unsatisfactory user experience. For virtual reality, a very common complaint from users is feeling of dizziness and motion sickness, caused by the specifics of the human perception and vestibular system, and heavily compounded by the possible low quality or high latency of the visual feed. For augmented reality, main cause of poor user experience is rather limited capabilities of the enduser device, leading, in turn, to poor performance of recognition systems. All this leads to a negative perception of the user's experience of interaction with such systems and, consequently, to the rejection of the use of technology.
- Lack of proper operational procedures when using the devices and poor legal base. For virtual reality systems, in particular in the educational process, it is critical to establish proper safety regulations and operational allowances to prevent unwanted sideeffects to the users. It is also very important to further study the issues of intellectual property and the potential violation of the boundaries of the user's personal life.
- The total cost of equipment and content for end users.

Together, these factors often lead to unreasonably high cost of equipment and content, leading, in many cases, to the practical impossibility of using these kinds of systems. Coordinating various approaches for designing and applying such systems on a conceptual level will facilitate the implementation of relevant solutions, which, in turn, can greatly increase interoperability, integration and convergence of all related systems and knowledge bases.

III. Suggested approach

This paper proposes a unified description for the process of designing virtual and augmented reality systems in the form of an ostis-system knowledge base [3]. As part of building the knowledge base and implementing a platform for developing systems in this subject area, the following stages are proposed:

- creation of a framework for semantic representation of the scene and user interaction;
- ystematization of the subject area, existing approaches and establishing links with related areas;
- development of a set of agents that implement the operational specifics of virtual or augmented reality systems.

Within the framework of the knowledge base, it is proposed to establish the following blocks:

- 3D representation of surrounding objects and scenes;
- description of the physical principles of operation and equipment specifications;
- description of the principles of user immersion when using a virtual reality system;
- technical vision methods and algorithms used in virtual and augmented reality systems;
- generation of images, object models and scenes in the user's visibility area;
- semantic representation description of threedimensional scenes and objects associated with the subject area

- description of scenarios of user interaction with the system.

Next, we will consider in more detail the basic principles of designing virtual and augmented reality scenarios.

IV. Semantic Representation Scenes and Interactions in Virtual Reality

The basis and a primary distinctive feature of all immersive systems is the creation of the effect of presence. To describe this concept and recreate the reality of the scene in human perception, paper [4] indicates the need to create three main illusions: illusion of place (inducing the feeling of being in a simulated place), illusion of realism (making sure the environment feels natural for the user) and illusion of impersonation (aligning user's virtual avatar to match their perception of self). These concepts are achieved by influencing human senses in a specific way. Primary methods are using panoramic stereoscopic displays (visual senses), surround sound (auditory senses) and tactile feedback (haptic senses), all of which combined also indirectly influence user's equilibrioception. At the same time, all designed interaction should be coordinated and take sensory-motor correspondence effect. Creating all the additional effects that a person perceives, even in a simple physical scene, is a very hard process that requires knowledge from many areas. At the same time, actual implementation for most of these effects can be inferred from context, semantic representation of the scene, and a proper knowledge bases in related subject areas.

Thus, using the semantic representation of the virtual reality scene, it is possible to establish action subject (user), description of the place and conditions for this action (may include an additional auditory and haptic feedback), as well as descriptions of the object of the action and all additional objects within the scene. Basic audio and tactile interaction can then be generated from this shared description. To support such a representation, it is necessary to create a knowledge base that includes basic actions and conditions, which will later be compared with the scene contents, user and object conditions.

Within the framework of OSTIS Technology, interaction and implementation of actions can be implemented by agents [5]. To design a virtual reality scene, two types of agents are proposed - generating agents and provisioning agents. Generating agents select an appropriate 3D representation and corresponding models according to the semantic content of the scene, determine and selecting the necessary auditory components, and select appropriate tactile feedback for all the interactions that can be performed by the user in this scene. Provisioning agents map this set of possible responses of the system, scene and user actions on the physical capabilities (configuration) of the end user equipment; this includes dynamic rendering resolution selection, ensuring proper reaction times for auditory and physical tactile interaction, and adjusting other scene generation parameters to maintain a balance between feed quality and perceived user latency.

To form a three-dimensional scene model, the semantic description of a three-dimensional scene given in [6] can be used, which consists of specifying individual objects, their absolute properties and referential relations. The use of referential relations can be effectively used in the process of rendering individual scenes.

For example, if the VR scene is of a person walking down a gravel path in a forest on a sunny day, then a 3D model of the surrounding forest can be generated from the database. In addition, the analysis of existing scenes can be used to generate background auditory effects (like the noise of foliage), and comparing actions and descriptions of the user and the environment can be used to generate reaction auditory effects (the sound of steps on a gravel road). Tactile interaction can be generated by user actions - for example, if the user picked up a pebble, it can be communicated using general vibration feedback in case of regular controller or a more detailed feedback in case of a haptic feedback glove. Therefore, actions are not coupled to specific equipment, it is possible to expand the capabilities of the equipment whenever necessary, and proper action feedback can be determined fully only by using the semantic description component.

Thus, using semantic representation of a virtual reality scene as part of scene generation pipeline contributes to producing a greater immersion effect due to using this representation as a single source of truth for all possible interactions, which guarantees effect and reaction consistency while allowing for extension and effective complexity management due to abstracting underlying interactions, feed modifications and end user equipment feedback.

V. Semantic Environment Representation in Augmented Reality

To design an augmented reality system, it is necessary to have a description of the physical world, a description of the informational world concept superimposed on it through augmented reality, and a description of their respective relationship. At the same time, the relationship should take technical vision system and object recognition method into account and ensure correct backprojection of the object model to display additional information on device screen in order to properly anchor virtual modelling space into the real world image and implement effects like object occlusion.

Augmented reality systems can use various types of sensors and, in turn, appropriate algorithms for orientation calculation and tracking. Existing local positioning and 3D reconstruction systems, as well as their representation within the OSTIS knowledge base, are described in