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 three-dimensional 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. Introduction

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 psycho-emotional 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 end-user 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 end-user 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 side-effects 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;
- systematization 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 three-dimensional 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 more detail in [6]. When designing augmented reality object tracking systems, all approaches for object tracking and scene anchoring can be divided into marker and markerless tracking [7].

To organize the operation of augmented reality system, four main types of agents have been identified:

- Device agents that process image from the camera or aggregate data from other types of sensors.
- Surface agents that are responsible for detecting and classifying surfaces for the purpose of anchoring, object placement and occlusion.
- Rendering agents that are responsible for visualization of 3D models.
- Application logic agents that are responsible for implementing user interaction scenarios and handling user input.

Later in this work, we also present examples of designing augmented reality systems based on semantic representation. When implementing concrete applications, in addition to designing the operation of the augmented reality system as an interface for the user, it is also necessary to describe relevant user scenarios. Designed scenarios can also be included into the ostis-system knowledge base and presented using an ontological description.

A. Application of augmented reality for technical device manuals

Technical device manuals and instructions are necessary elements of any type of hardware or equipment. Usage of augmented reality systems in the implementation of these kinds of manuals and instructions provides advantages in terms of clarity and visibility - for example, the user does not need to compare schemas, images or diagrams with the operating object in order to find the necessary structural elements and understand required operational, technical and maintenance procedures they need to perform. In addition, complex scenarios can be implemented: resolving non-standard situations can be presented in a sequence of steps, and intelligent visual analysis from the technical vision system can be used to spot and diagnose various issues. Many manufacturers are already implementing these types of applications, but at the moment all solutions in this area are limited to select products and distributed as narrowly-scoped standalone mobile applications. Using ontological representation of the technical specification and manual of the device [8] [9] makes it possible to design scenarios for personnel training, equipment diagnostics and repair. OSTIS Technology allows to extract an ontological description, associate it with a specific subject area and with the semantic representation of the augmented reality environment. As a result, ready-made user scenarios based on data received from the technical vision system are obtained. For the practical implementation of object recognition possible to use additional solutions or use the algorithms described in the knowledge base directly. The following entities can be described:

device
:= [subject of technical manuals and descriptions]
⇒ *includes**:
{
part of the device
⇒ *includes**:
{
device component
:= [A separate element responsible for specific functionality and having an additional description. Example: power button]
}
}

Each entity can be highly nested and must have a description and visualization provided in the database. A system component can be an integral part of the device. If it is necessary to project the corresponding 3D device models during AR environment modelling, these entities can be compared with entities within the framework of the semantic representation of 3D scenes and objects [6]:

3D object
:= [a set of points in space connected to each other and having a shared semantic representation]

3D composite object
:= [object in 3D representation that supports decomposition into separate individual objects]

part of an object in 3D
:= [a set of points in space belonging to some object in a three-dimensional representation, which can be distinguished by its geometric or semantic representation]

3D scene
:= [collection of several objects in three-dimensional representation and data about their features and relative position in space, including absolute and relative referential oriented and unoriented relations]

The following entity types are also introduced to describe user interaction:

equipment status
:= [a set of technical characteristics of equipment related to the specifics of its operation and determining its performance]

equipment operation

:= [a sequence of actions designed to achieve a specific user goal; examples of operations can be switching on, diagnostics, repair, installation, dismantling, transportation, etc.]

equipment action

:= [user interaction with one of the device components in a specified way; example of an action would be pressing a button, plugging into an outlet.]

Based on these entity types, an arbitrary custom scenario can be described. For example: turning on (operation) of new equipment (equipment status) is carried out by the following sequence of operations and actions: unpacking (can be an operation or action depending on the device), installation (operation), plugging in (action), pressing the power button (component). Device diagnostics or repairs can be described in the same way. The connection of each component to its visual display and semantic 3D representation allows to display descriptions of the stages aligned with the corresponding components using augmented reality, for example, on a smartphone screen.

As a practical implementation, an augmented reality training application based on markerless technology was developed with a technical manual of the oscilloscope for use during laboratory work. This application allows to visualise equipment operating instructions in AR, and also provides repair instructions using a mobile device that supports iOS 16. Device, surface and rendering agents implemented in this system are based on the ARKit SDK [10]. After the application is launched, the scene is initialized and a surface agent is used to find a surface matching the appearance of the device. When required surface is found, a WorldAnchor is set, and descriptions of individual device components that are present in the frame are displayed on the smartphone screen. While working with the application, the user can move in space, and appropriate component descriptions are constantly updated to correctly match location and spatial orientation of the device. Position recalculation is a computationally expensive task, so it is performed in 0.4 second increments. The user can also choose one of the operations to indicate how they want to interact with the device. In this case, the application enters instruction mode, and user can perform all the steps in order or view the entire sequence of actions.

B. Augmented reality quest

One of the most commonly used formats for augmented reality systems is gaming. Characteristics of this format can be used in educational systems, marketing campaigns, tours, and many other areas. One of the examples of designing an augmented reality applications presented in this paper is an augmented reality quest app - an educational and marketing application designed to familiarize applicants with an educational institution.

To create such an application, a set of scenes and interaction scenarios associated with them are described and added to ostis-system knowledge base, which will later be used as a data source.

Interaction implementation is based on a common location - some part of the space of the real world, which is described in terms of semantic scene representation and object relationships. These relationships provide a shared description of separate scenes and objects of these scenes in the real world. To design a scene in AR, additional relations are established between the objects of the real world and the virtual modelling space. Objects superimposed onto the scene are generated by the rendering agents, that use the same semantic description for the scene to work in a shared space. To perform visualisation of the resulting rendered scene, objects in the shared description and must be first located and recognized.

AR quest is formed as a sequence of the scenes and their respective descriptions. Each of these scenes also contains a description for a certain number of individual, which are, in turn, attached to the objects of the scene and can be presented to the user. An example of such a task is a set of questions about the observed object, a mini-game, an instruction to find a specific object, an instruction to move to the different location to change the scene. Application keeps track of task execution status using device agents that supply information about end user device spatial characteristics from device sensors. When the task is completed or scene change is detected, application may suggest new tasks.

To implement device, surface and rendering agents for proposed application, Unity platform and Vuforia software package [11] for AR were used. Developed system uses a combination of marker and markerless technology for tracking and searching for objects in a real scene. Application is designed to work under the Android OS. Managing content for describing scenes, objects of a real scene and supermposed objects of AR, as well as tasks for the user, can be performed using ontological representation of OSTIS knowledge base by implementing appropriate data mapping for translating ontological representation to the format used by corresponding platform APIs.

VI. Conclusion

The paper considers the possibilities of using OSTIS Technology for designing virtual and augmented reality systems, and discusses the possibility of using ontological descriptions of related subject areas as a basis for implementing these systems. For proposed concepts, descriptions of applications designed with the proposed approach are also provided.

The advantages of using OSTIS Technology for the presented tasks are:

- Introduction of a common concept and description system for different methods in a unified and consistent form.
- Support for convergence of VR and AR application design subject area with other applied subject areas, 3D scenes and environment modelling subject area, and computer vision subject area.
- Simplification of virtual reality systems development by selection of necessary scene elements according to scene semantic description to implement user interaction and achieve maximum user immersion.
- Ability to build a complex design technology using intelligent agents that rely on proposed description and utilise data about individual algorithms and user scenarios from the shared knowledge base.
- Ability to create integration tools for individual components, describe stages of various methods and build different types of internal representations in a unified way.

Using proposed approach enables to design dynamically expandable VR and AR systems that provide a greater immersion degree for the user, while using proposed semantic scene description for objects and their relations can be used to model various rendering and interaction scenarios without tightly coupling the implementation to specific hardware.

Building realistic scenes in virtual and augmented reality is a rather complicated process that must be able to achieve necessary degree of realism, while taking limitations on real-time scene rendering into account. Using OSTIS Technology to enable the possibility to use developments in related areas allows to form a new approach for developing less resource-intensive interaction scenarios.

Acknowledgment

The authors would like to thank the scientific team of Department of Intelligent Information Technologies of the Belarusian State University of Informatics and Radioelectronics for their help in the work and valuable comments.

References

[1] S. Bjork and J. Holopainen, *Patterns in Game Design*. Charles River Media, 2004.

[2] D. Saredakis, A. Szpak, B. Birckhead, H. A. D. Keage, A. Rizzo, and T. Loetscher, "Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis," *Frontiers in Human Neuroscience*, vol. 14, Mar. 2020. [Online]. Available: https://doi.org/10.3389/fnhum.2020.00096

[3] V. Golenkov, N. Gulyakina, and D. Shunkevich, *Otkrytaya tekhnologiya ontologicheskogo proektirovaniya, proizvodstva i ekspluatatsii semanticheski sovmestimykh gibridnykh intellektual'nykh komp'yuternykh system [Open technology for ontological design, production and operation of semantically compatible hybrid intelligent computer systems*, V. Golenkov, Ed. Minsk: Bestprint, 2021.

[4] M. Slater, "Immersion and the illusion of presence in virtual reality," *British Journal of Psychology*, vol. 109, no. 3, pp. 431–433, May 2018. [Online]. Available: https://doi.org/10.1111/bjop.12305

[5] D. Shunkevich, "Agent-oriented models, method and tools of compatible problem solvers development for intelligent systems," *Otkrytye semanticheskie tehnologii proektirovanija intellektual'nyh sistem [Open semantic technologies for intelligent systems]*, pp. 119–132, 2018.

[6] K. Halavataya and A. Halavaty, "3d representation of objects in new generation intelligent computer systems," in *Open semantic technologies for intelligent systems*. Minsk : BSUIR, nov 2022, pp. 251–260.

[7] J. C. P. Cheng, K. Chen, and W. Chen, "Comparison of marker-based and markerless AR: A case study of an indoor decoration system," in *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction*. Heriot-Watt University, Jul. 2017. [Online]. Available: https://doi.org/10.24928/jc3-2017/0231

[8] N. Dorodnykh, , O. Nikolaychuk, A. Yurin, and and, "Using ontological content patterns in knowledge base engineering for maintenance and repair of aviation equipment," *Ontology of Designing*, vol. 12, no. 2, pp. 158–171, Jul. 2022. [Online]. Available: https://doi.org/10.18287/2223-9537-2022-12-2-158-171

[9] A. Timoshenko, , A. Zuev, E. Mursalimov, V. Gribova, and A. Inzartsev, "Description and diagnosis of malfunctions in autonomous uninhabited underwater vehicles based on ontologies," *Ontology of Designing*, vol. 12, no. 3, pp. 310–324, Sep. 2022. [Online]. Available: https://doi.org/10.18287/2223-9537-2022-12-3-310-324

[10] P. Nowacki and M. Woda, "Capabilities of ARCore and ARKit platforms for AR/VR applications," in *Advances in Intelligent Systems and Computing*. Springer International Publishing, May 2019, pp. 358–370. [Online]. Available: https://doi.org/10.1007/978-3-030-19501-4_36

[11] M. Sarosa, A. Chalim, S. Suhari, Z. Sari, and H. B. Hakim, "Developing augmented reality based application for character education using unity with vuforia SDK," *Journal of Physics: Conference Series*, vol. 1375, no. 1, p. 012035, Nov. 2019. [Online]. Available: https://doi.org/10.1088/1742-6596/1375/1/012035

Проектирование пользовательского взаимодействия в иммерсивных системах нового поколения

Головатый А. И., Багай В. Д.,
Бернат Д. А., Головатая Е. А.

Данная статья посвящена рассмотрению вопросов проектирования интеллектуальных систем, использующих технологии дополненной и виртуальной реальности. Данные технологии основаны на использовании трехмерного представления окружающей реальной или спроектированной сцены, особенностях систем определения ориентации человека и объекта в пространстве, алгоритмах интеллектуального анализа и машинного обучения. Сложность разработки отдельных алгоритмов, необходимость учета специфики предметной области для каждой прикладной системы, отображение на основе трехмерных и других видов представлений вызывает необходимость разработки новых подходов и принципов проектирования. В статье основное внимание уделено возможностям Технологии OSTIS для использования в данных задачах, а также приводятся примеры построения системы технического описания оборудования и квеста в дополненной реальности.

Received 13.03.2023