A System for Hierarchical Browsing of Mixed Reality Content in Smart Spaces

Cristian I. Pamparău

Ștefan cel Mare University of Suceava
Suceava, România
pamparaucristian0@gmail.com

Abstract—**We present in this paper technical and engineering details of a wearable system for visualizing digital content. The digital content is overlaid onto the physical space by employing hierarchical structures for content organization. We implement analogies with the physical world adapting laws of physics, such as gravity, to digital content. To this end, we rely on concepts from the fields of Augmented Reality, Wearable Computing, Software Engineering, and Human-Computer Interaction to propose a new model for accessing digital content anchored in the physical reality. In our prototype, reality is augmented with containers of digital content represented graphically as semi-opaque spheres that float in the physical space around the user. We describe the properties of the spherical containers of digital content, such as their size and relative location in the physical environment, and their relationships to the user and the physical world. We also present interaction techniques based on gesture input that enable users to access the digital content contained by the virtual spheres. We present our technical implementation using the Microsoft holographic computer HoloLens and exemplify use case scenarios for our wearable prototype.**

Keywords—Mixed Reality, Augmented Reality, HoloLens, smartglasses.

I. Introduction

The idea of implementing a 3D Desktop in Augmented Reality is almost 20 years old, when Regenbrecht, Baratoff and Wagner [1] wanted to implement a tangible 3D desktop environment. At that time, the implementation mode assumed that the 2D standard was simply integrated into Augmented Reality, so that the 3D Desktop was practically a "display" of the 2D desktop, becoming tangible. Two years later, desktop-like systems developed for Virtual or Augmented Reality were already implemented by Mulder, Jansen and Rhijn [2]. Also, from the premise that the last 30 years have been dominated by the world of 2D desktops, Verdi, Nurmi and Holler [3] have also left, obviously, at that time the implementation was also in a way of designing only 2D digital content, in the augmented reality. In other words, subsequently, the "software visualization" term referred to the same thing and then attempted a 3D approach [4] for the visualization of digital content, starting from the premise that the visualization in 2D space was (is now) supersaturated, feeling the need to evolve to the next dimension, Augmented Reality being stopped in this situation.

II. Application Concept

In this paper we will present the 3D Desktop as a multitude of augmented realities represented by a series of semitransparent spheres, which contain within them diversified content, such as fruits, cars or boxes. In this way, the support application does not introduce the user into Augmented Reality to see a standard, transposed 2D desktop (menus, lists, buttons, etc.) but only visual components that are totally different from a regular Desktop. The size of the spheres, but also the distance from the sphere to the earth reflects the size of the digital content.

The basic idea is that a sphere represents a directory from the standard 2D desktop, as we know it, and the fact that it contains objects means that the directory is one that contains, in turn, other directories or files. Thus, the size of the memory occupied by a directory is given by the sum of the size of each element inside. In this sense, the size of the sphere will be influenced by the amount of memory occupied by the elements inside. The fact that a larger sphere is attracted to more ground, means that the 3D director occupies a large space in memory. Given that the augmented reality is intended to be superimposed on the physical one, it is important that the built system respects and implements laws of the physical world, such as gravity, and the human mind already considers an axiom that anybody in the physical world that has weight, is attracted to the ground. The idea is, therefore, to convert the digital information that we also know from the 2D world into properties of the physical world. So, returning to our app, the visualization system will contain a series of 3D directories located in augmented, virtual space, superimposed over the physical space. We call this first view the level 0 hierarchy of digital content.

In terms of data manipulation, the digital content must be interactive and accessible for users, considering that an AR Desktop "is used as the interaction spaces" [11]. Once these have been accessed, virtually the visualization system will display the contents of the respective sphere, that is, access to level 1 of the digital content, the next level in the hierarchy. It is important to keep in mind that the sphere, visualized at level 0, will contain only one object, representing that that virtual directory includes only objects of that type. In this sense, once

accessed level 1, more information will be visible, more objects of the respective type, their number directly influencing the size of the sphere. Finally, the system must necessarily support a way to return to the previous level, this happening at the access of any file in the content of the sphere.

At this point, if we think of a standard 2D desktop, it is defined by a series of operations, the first and most important being the visualization. Secondly, accessing the information hierarchy defines the system otherwise it would be just a folder containing a series of files. If we consider these two stages as defining for such a concept, we can say that the support application for this paper constitutes a prototype for an interactive 3D desktop, consisting of several augmented realities that overlap over the same physical reality.

III. IMPLEMENTATION DETAILS

In order to implement the concept described above, we used Microsoft HoloLens technology which is "the first fully self-contained holographic computer running Windows 10" [5]. This involves a design stage and a programming, definition and modeling stage of the previously designed scenes.

For the design stage, the Unity3D gaming engine was used, and the augmentation of the realities was achieved with the help of the C# programming language, Visual Studio as IDE. The interaction with the objects from the implemented reality scene "is built on gaze to target and gesture or voice to act upon whatever element has been targeted." [6]. In this regard, accessing the next level of hierarchy of digital content is achieved primarily by locating the object to be accessed. The application implemented the first form of input, Gaze.

According to official documentation, "they gaze at you where the user is looking in the world and lets you determine their intent" [7] as in the real world, when you want to interact with an object, you first look at it. The support application for the present work has implemented this by modifying the cursor appearance when the user has the target positioned on an object. Those two types of cursor appearance can be visualized in Fig. 2.

Once an object has been located, identified, it can be accessed by performing Air Tap technology gesture. This is "a tapping gesture with the hand held upright, similar to a mouse click or select" [8]. Most of the time, users perceive this gesture and associate it with the "click" event from a standard 2D desktop. An important aspect to note about this gesture is that it must be executed so that it is visible to the HoloLens camera, so that the device can recognize this gesture. Thus, the object can be targeted, the gesture executed correctly, but the camera does not surprise this gesture, and therefore the expected effect will not happen. On the other hand, it is as possible that the object is targeted, the camera suppresses the gesture, but it is executed incorrectly. On this topic, Microsoft states that "Users need to be trained in this area of recognition both for success of action and for their own comfort (many users will initially assume that the gesture frame must be within their view through HoloLens, and hold their arms up uncomfortably in order to interact)" [9].

```
1   public void OnInputClicked(InputClickedEventData eventData) {
2     GameObject[] objects = SceneManager.GetActiveScene().GetRootGameObjects();
3     if (this.gameObject.name.Contains("Sphere")) {
4       destroyPrefabsByName("Sphere", objects);
5       if (this.gameObject.name.Contains("Box")) {
6         loadPrefabsByName("Box");
7       } else if (this.gameObject.name.Contains("ford")) {
8         loadPrefabsByName("ford_GT");
9       } else if (this.gameObject.name.Contains("Fruit")) {
10        loadPrefabsByName("Fruits");
11      }
12    } else {
13      destroyPrefabsByName(this.gameObject.name, objects);
14      SpawnItems.Instance.SpawnManyItems();
15    }
16  }
```
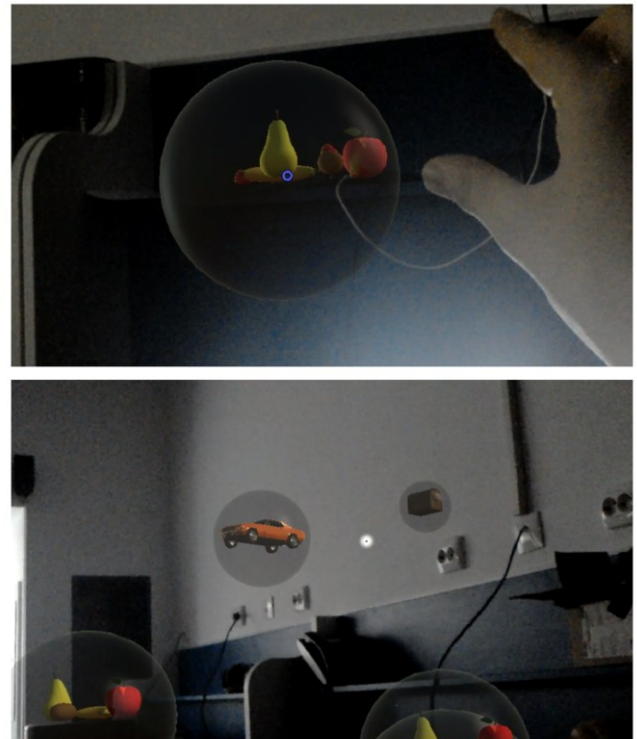
Fig. 1. The Behavior of AirTap Gesture

Fig. 2. The cursor appearance. *Top:* the cursor appearance when a object is gazed, so that the AirTap gesture can be made. *Bottom:* the cursor appearance when no object is gazed. Please note that when an object is not gazed, the correct execution of AirTap gesture will have no effect.

For the implementation of the gestures, it was necessary to insert in the design stage a package that ensures the implementation of event-related procedures, such as OnInputClicked. Thus, once generated the behavior of the event, the script in which it is described, will be attached to the objects that want to have the respective behavior, the script being part of the objects of the scene. The tools used are those that are part of the MixedReality-Toolkit [10]. The behavior of the OnInputClicked function is shown below.

If we analyze the code from Fig. 1, we will notice the behavior described in the explanation of the concept. Thus, if the reached object is a sphere (it contains the word "Sphere" in its name) all objects from the current scene containing the word "Sphere" will be destroyed (as we said, in the level 0 there are just spheres/directories, so the scene will be empty in this first

step). In the next step we will identify the type of sphere that was touched. For instance, if it was a sphere of boxes (the initial name contains "Box"), the previous cleared scene will be redecorated with a random number of boxes (loadPrefabsByName("Box")) or, the next level will be accessed. If the name of the object reached does not contain the expression "Sphere", it means that the current level is level 1 and all the objects in the scene will be destroyed and the user will be returned to primary level by calling the SpawnManyItems() method from the SpawnItems script instance.

IV. OBSERVATIONS FROM A PUBLIC PRESENTATION

In a public exhibition within a Shopping Center in Bucharest, an exhibition on the theme of Augmented Reality, the application was tested and discussed by a multitude of visitors, aged between 10 and 60 years. Please note that this was not a controlled experiment, it was just a presentation of the application to a wider audience in order to request feedback that which i used it to draw the following conclusions. The reactions of the users regarding the technology and the concept implemented were different, based on different life experiences, their culture and education, but mainly on their contact with the technology.

As for the application, as we anticipated, it was difficult for users to agree that what they saw would be a desktop. Most expected to actually see a projection of a 2D Desktop, lists, buttons, menus, etc. This speaks clearly to the fact that the move to another kind of desktop will be done gradually, as people will be open to thinking or imagining their digital content in a different way than the classic one.

At first impression, visitors were interested to test the application, but the impediments of lack of experience with technology appeared immediately. Thus, one of the impediments was that, after learning to properly execute the AirTap gesture, the next level of hierarchy was inaccessible. This was due to the fact that the users focused more on the correctness of the gesture and its positioning in front of the camera, but the target was not achieved (gaze). Basically, they did not interact with any object, which, of course, had no effect. Another series of users reacted in the opposite direction: they focused on the object, but either they did not make the correct gesture from the first, or they did it correctly, but not in front of the cameras or in the field of view's cameras. Below, a series of images from the experiment are shown in the Fig. 3.

Regarding Fig. 3, at the exhibition, children and teenagers were interested in exploring the device and technology first. Even though it was a new technology that they had not come in contact with, exploring the device and the application, later, it was considered a pleasure. The adults, on the other hand, considering testing the application a challenge, preferred a verbal and conceptual presentation of the application, very few of them being curious to test the application. Even if the way they came in contact with the application/concept was very different, the conclusions and the way they perceived the experiment was the same: it takes time to be familiar with AR/MR applications, it requires a familiarity with the technologies and virtual environments so that the presentation

Fig. 3. Photographs captured during an open exhibition to the public, showing visitors trying out our application for HoloLens.

of any new concept is deductible, the semantics of the applications became easy to understand and do not require too many explanations on the question „What is this [application]?". In fact, I think that the problem is not understanding the concept/application, but understanding the definitions for AR/MR realities.

In other words, it is unrealistic to expect AR/MR applications to be easily understood since the AR/MR domain has not yet been well clarified and defined at a comprehensive level.

If we look in the scientific literature, we will notice that the definitions for these realities are divided. The first definition came from Azuma [13] when he states that Augmented Reality is defined by a system that is three dimensional, is interactive in real time and combines real and virtual environment. Milgram [14] defined Mixed Reality as an environment in which real world and virtual objects are presented together within a single display. Later, Mann et al. [15] defined MR as an experience in the same time of both real and virtual world, a mix of real and virtual worlds in various proportions. More than that, same authors believed that is "highly unrealistic to expect [definition] to appear in the future" (p. 12). Based on these statements and definitions, what Mann told about mixed reality, Azuma states about augmented reality, so that the distinction between these realities is hard to be made. After years of research on this topic, Speicher et al [16], referring to a definition of mixed reality and trying to outline a concrete idea about the field, identified a general definition: "it depends" (p. 12), arguing that MR can be represented as "many things and its understanding is always based on one's context" (p. 12). So, how can we expect users to understand with no explanations the content of AR/MR applications since there is no concrete definition, generally valid and accepted for these realities?

The solution and answer to this question would be to assist and guide users when they contact this type of application, as you can see in Fig. 3.

V. CONCLUSION AND FUTURE WORK

This paper proposed a hierarchical way of presenting digital content in an AR environment. Those two levels of information are accessed by performing AirTap gesture, but the future work should deal with at least three levels of information. In order to do that, using just AirTap gesture for interaction will be inappropriate. For example, if the user will be on the middle level, by accessing a sphere, will access the last level and how will return on the primary level? In this way, Bloom gesture [12] needs to be implemented. The Bloom gesture supported only with the HoloLens 1st generation is, according to the official documentation „a symbolic gesture mimicking the flower blossom".

Another future directive for the current project should be supporting more Desktop operations, such as merging two spheres or moving a sphere in the space; also, another idea could refer to sharing the same augmented space to multiple users form the same location.

REFERENCES

[1] H. Regenbrecht , G. Baratoff and M. Wagner, "A tangible AR desktop environement", in Computers & Graphics, Volume 25, Issue 5, pp. 755-763, October 2001.

[2] J Mulder, J. Jansen and A. Rhijn (2003, May 22-23), "An affordable optical head tracking system for desktop VR/AR systems", in Proceedings of the workshop on Virtual environments 2003 (EGVE '03). Association for Computing Machinery, New York, NY, USA, pp. 215-223, May 2003.

[3] S. Di Verdi, D. Nurmi and T. Hollerer, "ARWin - a desktop augmented reality Window Manager", in The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, Proceedings., Tokyo, Japan, 2003, pp. 298-299.

[4] A. R. Teyseyre and M. R. Campo, "An Overview of 3D Software Visualization," in IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 1, pp. 87-105, Jan.-Feb. 2009.

[5] Microsoft HoloLens (2019). Microsoft HoloLens. [Online] Available: https://docs.microsoft.com/en-us/hololens/ [Accessed: March 2020]

[6] Microsoft HoloLens (2019). Gestures. [Online] Available: https://docs.microsoft.com/en-us/windows/mixed-reality/gestures/ [Accessed: March 2020]

[7] Microsoft HoloLens (2019). Gaze. [Online] Available: https://docs.microsoft.com/en-us/windows/mixed-reality/gaze/ [Accessed: March 2020]

[8] Microsoft HoloLens (2019). The two core gestures of HoloLens: AirTap. [Online] Available: https://docs.microsoft.com/en-us/windows/mixed-reality/gestures#air-tap/ [Acessed: March 2020]

[9] Microsoft HoloLens (2019). Gesture Frame. [Online] Available: https://docs.microsoft.com/en-us/windows/mixed-reality/gestures#gesture-frame/ [Acessed: March 2020]

[10] Microsoft HoloLens (2019). MixedRealityToolkit-Unity. [Online] Available: https://github.com/Microsoft/MixedRealityToolkit-Unity/ [Acessed: March 2020]

[11] S. Jeon, H. Shim and G. J. Kim. "Viewpoint Usability for Desktop Augmented Reality", in International Journal of Virtual Reality, IPI Press, 2015, 5 (3), pp.33-39.

[12] Microsoft HoloLens(2020). Bloom Gesture. [Online] Available: https://docs.microsoft.com/en-us/windows/mixed-reality/system-gesture#bloom/ [Acessed: March 2020]

[13] Azuma, R.T., "A Survey of Augmented Reality. Presence", im Teleoper. Virtual Environ. 6(4), pp. 355–385, 1997.

[14] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays", in IEICE Transactions on Information and Systems E77-D(12), pp. 1321–1329, 1994.

[15] S. Mann, T. Furness, Y. Yuan, J. Iorio, Z. Wang, "All Reality: Virtual, Augmented, Mixed (X), Mediated (X, Y), and Multimediated Reality", CoRR abs/1804.08386, 2018.

[16] M. Speicher, B.D. Hall, M. Nebeling, "What is Mixed Reality?", in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19. Association for Computing Machinery, New York, NY, USA, paper 537, pp. 1–15,2019.