MARKER-BASED FINGER GESTURE INTERACTION IN MOBILE AUGMENTED REALITY

Loubna Ahmed, Doaa Hegazy, Salma Hamdy, Taha Elarif

Abstract: This paper proposes a mobile AR application that allows the user to perform 2D interactions with the 3D virtual objects. The user is supposed to hold the mobile with his left hand and interact using the other one. In addition, the user is supposed to attach two colored markers (stickers) to his fingers; blue for the thumb and green for the index. User studies (experiments) were conducted to test the different interaction types; translation, scaling and rotation. The application run on a Samsung Note 5 device with Android 7 as OS. Our results were based on the performance (completion) time per each task per each participant in addition to a subjective questionnaire that was answered by the participants after finishing the user studies. According to the results, it was found that this approach had a delay which implies low performance and users faced a slight difficulty in accomplishing all tasks, yet this approach proved to be engaging and fun.

***Keywords**: Mobile Augmented Reality, Marker-based Interaction, Color Detection.*

***ITHEA Keywords**: H.5 INFORMATION INTERFACES AND PRESENTATION, H.5.1 Multimedia Information Systems.*

Introduction

Over the past years, Augmented Reality (AR) has evolved and one of its evolutions is mobile AR. A key point of mobile AR is being reactive, which imposes real-time constraints. Interaction techniques focus on allowing the users to interact with the emerging virtual object and are considered the basis for having a successful AR system.

The interaction techniques can be categorized into tangible and intangible, adopting a classification introduced by Bai H. et al. [Bai et al, 2012]. A survey about tangible and intangible techniques was presented in [Ahmed et al, 2015]. Tangible interaction techniques refer to the type of interaction where the user physically touches something, whether a mobile screen (touch-based) or a keypad (device-based) [Bai et al, 2012]. On the other hand, intangible techniques refer to the systems where the user has no physical connection with the environment, such as midair gestures. The interactions that can be implemented with the virtual objects are the transformations; translation, rotation and scaling.

One of the intangible techniques is the finger-based gesture interaction. Finger gesture interaction techniques can be either 2D or 3D; hence the virtual object can be transformed in 2D or 3D. They rely on detecting the user’s hands and(or) fingers. For fingers detection, finger tips can have markers attached (marker-based) or marker less. 3D interaction needs an extra camera; for example, Kinect or Prime Sense to capture the fingers in 3D. While 2D can only rely on the mobile device in-built camera.

In our system, we implemented marker-based intangible interaction with the virtual objects. There are various applications where this approach is needed; such as games and education. Educational applications let students interact and get engaged in what they learn by making the content visible and interactive. For example, in history, students can interact with historical sites as if it is alive. Also, in science like physics and chemistry where everything is invisible; for example, molecules and chemical reactions can be visible and interactive. In 2018, authors in [Syahputra et al, 2018] presented an application for offering information and experience about the endangered animals in Indonesia through virtual objects of those animals.

Our motivation is to provide a 2D midair marker-based finger interaction with the 3D virtual objects and to provide the user a feedback upon interaction. So, the key points for our research were the detection of the markers attached to finger tips, handling the different interactions. The types of interactions handled are selecting an object (with one or two fingers), translation, scaling and rotation.

In this paper, we begin with the related work in section 2. The proposed system and how the system provides the user with visual feedback is illustrated in section 3. Section 4 discusses the user studies conducted and the results are stated in section 5. Finally, the conclusion can be found in section 6.

Related Work

Intangible interaction refers to the type of interaction that relies on the physical separation of the user from the device; like midair gesture, speech, etc. The interaction is mapped onto input parameters to control virtual content [Bai et al, 2012]. Intangible techniques can be classified into either marker based where markers are attached to the fingers or marker-less.

Authors in [Hürst & Van Wezel, 2013] proposed a marker-based interaction with both virtual and real objects using 2D input and a single camera (the camera of the device). Two experiments were made: one for virtual objects floating in midair, and another when the objects have a connection to physical ones. Only 2D interactions are handled in this paper as 3D tracking with one camera on a mobile has limitations. In the first experiment, a sticker is attached to the fingertip. The second experiment, a green and a red marker are attached to the user’s thumb and index finger, respectively. A bounding box is generated around both markers and the virtual object such that interaction is detected upon overlapping or touching between these bounding boxes. Translation, scaling and rotation were implemented. The object appeared as selected or not, hence alleviating problems such as lack of haptic feedback. It was found that the touch-based concept has the best performance while the finger-based is ranked to be the most fun. The results show that translation using one or two fingers worked well but the users preferred using two fingers as it is more natural. For scaling, within one interaction type, both midair and on-board operations are almost the same, but they differed with respect to accuracy. Finally, for rotation, using one finger and two fingers differed significantly in terms of average time while rotation in midair and on board differed slightly.

In 2008 [Seo et al, 2008], Seo et al. proposed a one-handed interaction technique where virtual objects are augmented on the palm of the user’s free hand. The author introduced both visual and tactile interactions. The visual interaction technique does not need any external visual markers or tags, as interaction is done by motions of the hand such as opening and closing of the palm causing the virtual object to respond to the pose changes. On the other hand, the tactile interaction is obtained through receiving feedback from the virtual object. The authors presented an application for this model, where the user interacts with virtual pets.

Choi et al. proposed in 2011 a bare-hand-based AR interface for mobile phones [Choi et al, 2011]. This interface is similar to what Seo. et al. proposed in [Seo et al, 2008] but with more accurate estimate of all possible palm poses. Moreover, their proposed methodology is less time consuming.

A finger gesture-based intangible technique based on midair gestures is introduced in [Bai et al, 2012]. They implemented a finger tracking system (C++ in Android NDK). The hand is segmented from the background if the hand area detected exceeded a certain threshold value. Then the prominent fingertips are marked. In this technique, skin color detector working in HSV color space are applied followed by the distance transform. The fingertips are then identified based on the curvature-based contour point sampling and elliptical fitting method used in Handy AR [Lee & Hollerer, 2007]. This implementation works under stable light condition and with an assumption that the hand is placed in front of the user’s face, and during the gesture interaction, the finger is always visible. It was found that the gesture-based concepts took more than twice the time of the freeze-view and the free-view touch.

Gao [Gao, 2013] presented two 3D gesture-based interaction methods for mobile AR. Both methods support 3D interaction by using depth camera to obtain 3D coordinates of the users’ fingertips along with the virtual objects. Moreover, a touch-based method was introduced in this thesis and compared to the natural gesture-based methods. The first technique, Gesture-Based Interaction using Client-Server Framework, consists of a PC desktop as a server, a Kinect depth camera, a tablet as a client, and a marker. The Kinect camera sends the RGB and depth images to the PC server which combines them to get the depth information of the images. The fingertip 3D coordinates are detected by this calibrated RGB and depth output and sent to the mobile client for mobile AR 3D interaction via a wireless connection. This framework implements three atomic gestures: translation, rotation and scaling of objects. The second technique, Gesture-based interaction using a tablet, consists of a tablet and a Primesense depth sensor connected using a USB cable. It just combines the PC server system and the mobile client system together into a tablet. RGB and depth images are acquired from the depth sensor and combined to support a pixel to pixel mapping for the system. The fingertip coordinates are calculated based on the combined images and a full 3D manipulation is supported for the users by using the received fingertip coordinates. This method implements pinch-like gestures for selecting objects which simulates the real life for grabbing real objects. This is achieved by comparing the midpoint between the two finger tips; thumb and index with the center of the object. The authors presented their results comparing the gesture-based interaction with the touch-based; stating that on average, 2D-touch based interaction performed better and faster than 3D gesture-based interaction. In addition, according to the subjective questionnaire; it was found that, in general, gesture-based is more fun and engaging than touch-based. However, touch-based has proven to be much easier and less stressful. To sum up, the results were not as the authors expected as most users preferred touch-based to gesture-based. They analyzed the reasons as follows; there is no physical feedback to the user from the objects on selection, the user’s fingers are always covered by the virtual objects overlaid on the video frames. Another reason is the problem of losing the fingers position when the user’s free hand gets out of the video frame accidentally. Finally, a problem arises when using two fingers for translation or scaling; as the distance between the two fingers can be greater or smaller than the virtual object size, also while rotating the object one finger can cover the other in certain angles. All these problems need to be considered and handled in the future.

Baldauf et al. [Baldauf et al, 2011] developed a visual marker less fingertip detection engine to detect 3D objects along with several use cases for using this engine in interaction. One of these use cases is the Virtual Object

Interaction, where the virtual objects can be manipulated through selection by pointing, grabbing or dragging and dropping by pinch like gesture using the thumb and index fingers. They implemented a fingertip detection methodology.

In 2013, Chun and Höllerer [Chun & Hollerer, 2013] introduced a methodology for marker-less real-time handling interaction of users with the virtual objects appearing on the mobile phone screen. In this paper, three gestures are handled; translation, scaling and adjusting the transparency of the object as the user’s hand come closer or moves away from it. Those gestures were specifically chosen to optimize the learn ability as they are similar to touch screen gestures. Authors assumed the hand will be around the AR marker (virtual object) hence, a search window of a size three times the marker area is chosen to save computational costs. The search window is divided to 4x4 grids to track hand movement within the area by computing the percentage of skin-color pixels P(skinn). Two different interaction modes are implemented: one is discrete event detection where a threshold is set manually for each cell. The second is continuous value adjustment where they recorded how much the hand occludes that grid. In this case, each cell is not compared to a threshold, but the amount of hand occlusion in each cell, is used to change a value dynamically, for instance opacity.

In 2017, authors in [Syahputra et al, 2017] presented a finger detection methodology in the historical domain. The main objective of this system is to display 3D objects based on the users’ interaction through detecting the number of fingers that the user use. Their system obtains the scene through a web camera, identifies the hand structure and detects the number of fingers by convex hull and convexity defects.

The Body

Our proposed framework as shown in Figure 1 works as follows. Initially, a bounding box around the virtual object is created. Then, the fingertips markers are detected and mapped to the same space of the object’s bounding box. At this point, the system checks whether a collision took place between the fingertips and the object or not. In case of collision, which implies a successful selection, the system detects the type of interaction performed by the user. According to the detected interaction type, the system responds and transforms the virtual object. Finally, the object's bounding box is updated according to the transformation done.

AR tracking has been built over the natural feature tracking library (Vuforia 6.2) for Android mobile applications. The experimental device used was Samsung Note 5 with Android 7 running on it. Markers' detection on fingertips was implemented by OpenCV for Android as discussed in the following section. Also, OpenGL for Android was used.

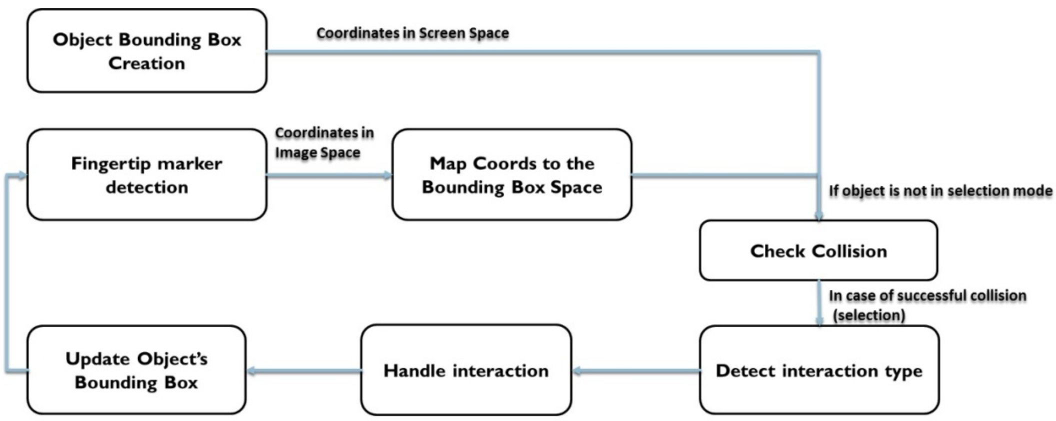


Figure 1. Proposed System

Finger tips markers detection

In our approach, markers are attached to the fingers; green and blue markers for the index and thumb respectively. These markers will help easily detect and track the finger tips. The colored marker detection was implemented using OpenCV library with Android and Vuforia library. The purpose of this step is to detect the position and size of the colored markers attached to the fingers.

One of the limitations of our approach is the position of the mobile phone device with respect to the AR marker; the mobile device is assumed to be on the left side of the marker.

Marker detection is conducted by capturing the scene (current frame) from the camera in RGB (Red, Green and Blue) format. The image is then converted into HSV (hue, saturation and value) because it is much easier to threshold images in HSV rather than RGB. Thresholding then takes place to detect both blue and green colors. A morphological opening followed by morphological closing are performed. The opening is to remove the small objects from the foreground. While the closing is to fill the small holes in the foreground. By this step, the largest contour representing the intended finger is surrounded by a rectangle. This rectangle is obtained in the coordinate system of the image (frame). As a result, some conversions on this bounding rectangle’s vertices take place to match the same coordinate system of the virtual object, which will be illustrated in the next section.

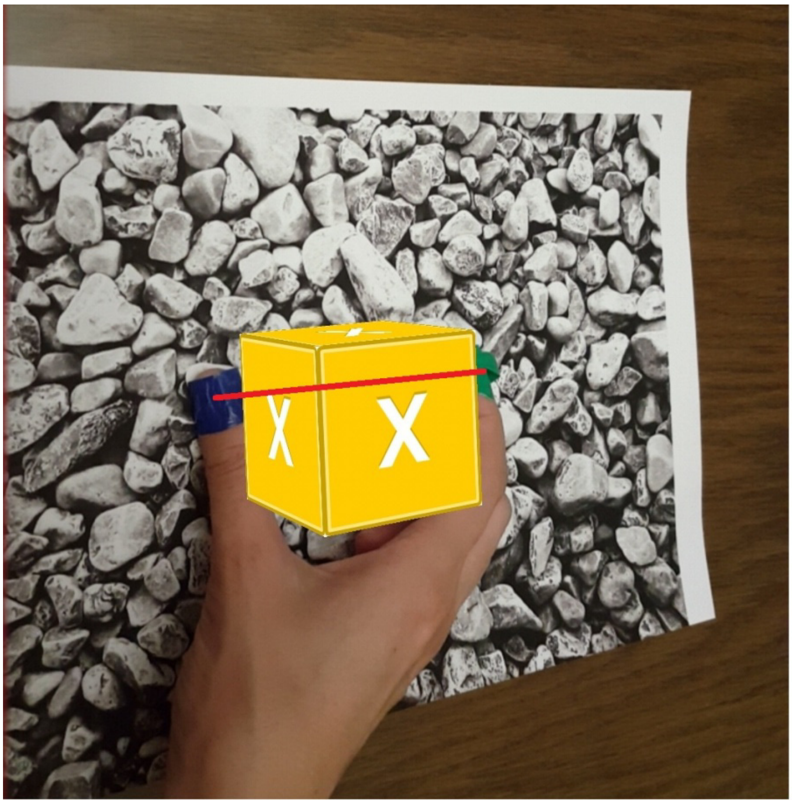
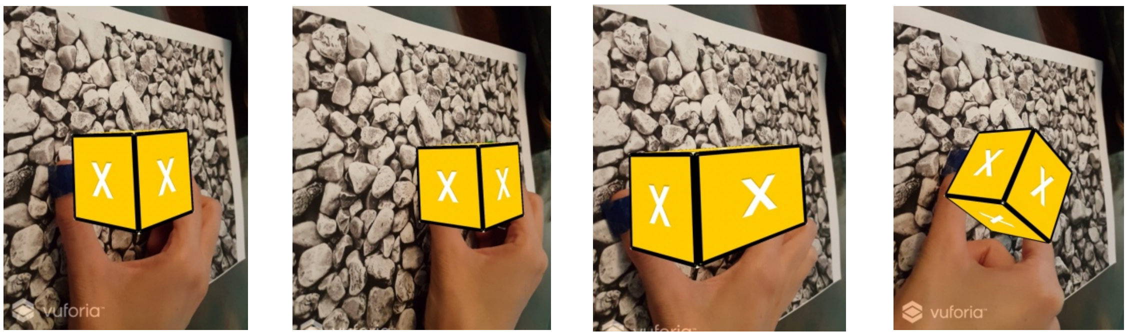


Figure 2. The selection mechanism



(a) (b) (c) (d)

Figure 3. (a) shows the object selection by the two fingers. (b) shows the object after being translated to the right direction. (c) shows the object after scaling (zooming in). (d) shows the object after rotation in clockwise direction.

3D Object Bounding Box Calculation

After detecting the finger tips, a bounding rectangle is created around the intended virtual object. The purpose of this step is to get the object's position and size.

Initially, the position is calculated through the Vuforia library, by mapping the point (0,0,0) which is the center of the target plane to a 3D camera point. Then, this 3D camera point is projected to screen point. In the next steps, the position will be maintained upon interactions.

Coordinate System Mapping

As illustrated previously, the coordinate system of the markers is different from that of the object's bounding box. The difference is not only in the values of the scaling, but also in the orientation. Thus, the following calculations (1), (2), (3), (4) and (5) have taken place on the markers' coordinate system. First, the following conversions are applied on both the x and y coordinates of the both the top left and bottom right vertices.

$$x' = \frac{(y * width_{screen})}{height_{image}} \tag{1}$$

$$y' = \left(height_{image} - x\right) * \frac{height_{screen}}{width_{image}} \tag{2}$$

Next, the center point and the size are calculated as follows,

$$center\ point = (\frac{{x_{tl}}' + {x_{br}}'}{2}, \frac{y'_{tl} + {y_{br}}'}{2}) \tag{3}$$

$$width = x'_{tl} - x'_{br} \tag{4}$$

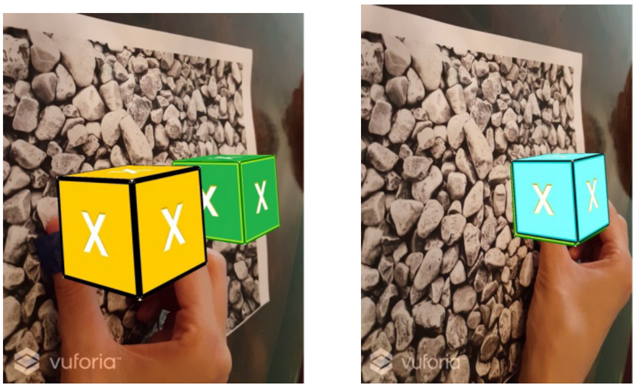
$$height = y'_{br} - {y_{tl}}' \tag{5}$$

where *tl* stands for the top left coordinate and *br* stands for bottom right.

Interaction Types

The object is successfully selected when the center of the object lies between the centers of both index and thumb fingers. In other words, the midpoint between the centers of both fingers lies within a threshold near the center of the object, as shown in Figure 2.

On the first successful selection of the object, the object's state becomes selected (Figure 3(a)) and the system is ready to detect the type of interaction performed; translation, scaling or rotation.



(a) *(b)*

Figure 4. The translation tasks. (a) shows the manipulated cube in yellow and shaded in black and the target cube in green. (b) shows the manipulated cube after matching the target and colored in aqua

Table 1. Questionnaire about the proposed system

The interface was:
Q1. Easy to learn
Q2. Easy to use
Q3. Natural
Q4. NOT mentally stressful
Q5. NOT physically stressful
Q6. With fun and engagement

Translation (Figure 3(b)) of the object is done when both fingers are moving in any direction together, such that if the index's displacement in the positive x direction is increasing, the thumb's displacement in the positive x direction is to be also increasing.

Rotation is triggered when the movement of the one finger is opposite to the other finger as shown in Figure 3(d). For example, in case of clockwise rotation, the thumb is going upwards in the negative y direction while the index is going downwards in the positive y direction. The angle of rotation is calculated as follows

$$angle1 = atan2\,(y_{index\,previous} - y_{thumb\,previous},\quad x_{index\,previous} - x_{thumb\,previous}) \tag{6}$$

$$angle2 = atan2\,(y_{index\,current} - y_{thumb\,current},\quad x_{index\,current} - x_{thumb\,current}) \tag{7}$$

$$angle = (\,angle1\ in\ degrees)\ \%\ 360 - (angle2\ in\ degrees)\ \%\ 360 \tag{8}$$

On the other hand, as shown in Figure 3(c), scaling is done when the two fingers are either going towards each other; the distance in between both the two fingers' centers is decreasing in case of zooming out or far away from each other; the distance between both the two fingers' centers is increasing in case of zooming in. The scaling value is computed by the difference distance between the two fingers either increasing or decreasing.

Feedback

One of the limitations in many proposed approaches was the lack of haptic feedback. In our research, we intended to implement a simple yet influential feedback; changing the color of the borders of the object upon successful selection of the virtual object.

In our approach, the virtual object becomes shaded in black color upon successful selection (Figure 3(a)). Hence, the user can be informed that the object is ready to be transformed. In addition, when the object is no more being held by the user’s fingers, it returns to its original color; pure yellow.

Implementation of results

The goal of our experiments is to measure the time taken by users for all tasks in the experiment and the user satisfaction based on a subjective questionnaire that was answered by the users after performing all tasks in the experiments.

To investigate our approach, we implemented a set of user studies to test the performance. In addition, users were given a questionnaire to answer regarding their feedback about the approach.

To perform the different user studies, we recruited participants. Their age range varied from 10 years to 57 years. All of them were right-handed. Also, all participants had experience with smart phones. All participants did not experience Augmented Reality interfaces on mobile devices. Each participant had to perform five tasks. In each task, the user had to transform the cube (figure 4(a)) to reach and match a target cube as shown in the figure 4(b). Participants were given an introduction about the system, how to use it and how the user studies work. Moreover, participants were asked to perform the tasks as fast and accurate as possible.

The first two tasks, the user had to translate the cube, once in the positive x direction (right direction), and the other time in the negative y direction (downwards). The next user study was rotating the virtual object on clockwise. The forth study, was scaling the virtual manipulated cube to a certain factor. The last task was a hybrid between scaling and translation. The system automatically calculated the performance of each task for each user in terms of task completion. Besides, the questionnaire was made of six questions [Bai et al, 2012] to be answered by the user as shown in the Table 1.

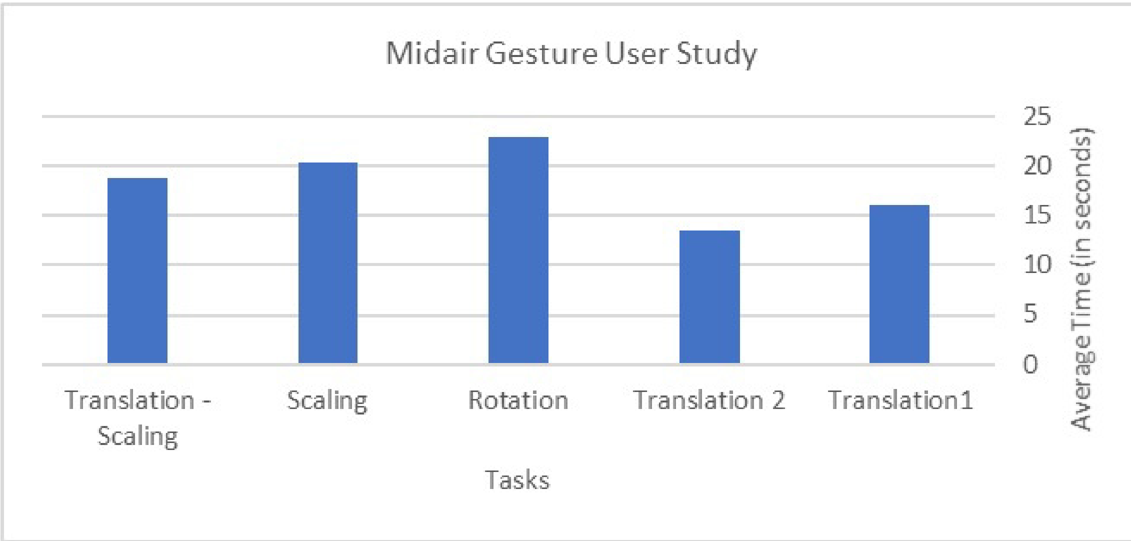


Figure 5. Average Completion Time for each task in Midair Gesture Approach

Results

Figure 5 shows the average time (in seconds) taken by the participants to complete the tasks. It was found that the translation tasks had the best performance with 17 seconds for the first task and 14 seconds for the second task and easiest for users based on their answers. Scaling comes as the next in performance with 22 seconds. While, rotation had the lowest completion time with 24 seconds and users found it to be the most difficult to accomplish. Hybrid tasks, in average, consumed much more time than those with only one task with average 20 tasks.

As per the subjective questionnaire, Figure 6 illustrates its results based on the answers of the users. It was found that most users did not think that this approach was not as easy to use as it was expected. On the other hand, it proved to provide users with fun and engagement. In addition, it was found that users agreed that the proposed approach was neither mentally nor physically stressful.

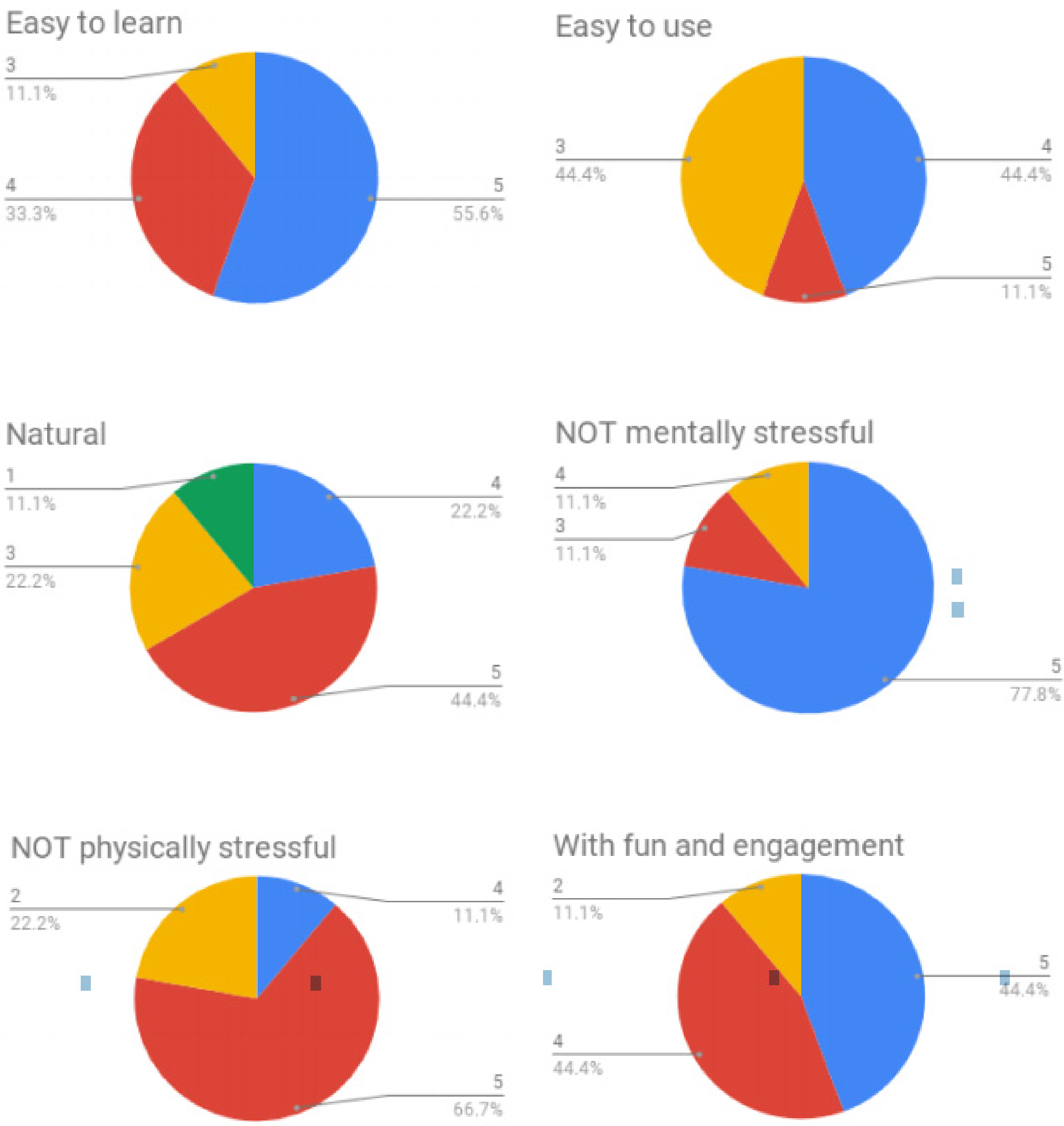


Figure 6. Results of Subjective Questionnaires

Conclusion

In this paper we presented an intangible interaction approach were the user manipulates virtual objects augmented on the real world with his fingertips. Our proposed approach let the user selects (holds) the object using both index and thumb fingers, then he can translate, scale or rotate the object using both fingers. For the fingertips detection, colored markers were attached to the fingertips of the users, green and blue for index and thumb respectively. After detecting the positions of the colored markers, the position of virtual object was determined. Then, the system checked if there was a collision between the fingers and the virtual object. On the first collision, the object was marked as selected (held) and the system was ready by that time to detect the type of manipulation performed; translation, scaling or rotation. Consequently, the object was correctly transformed according to the manipulation performed.

One of the limitations of the proposed system is the performance; the process of colored markers detection consumes time resulting in a slight delay. Also, the transformations were not as natural and accurate as we assumed. Thus, the accuracy of the object transformations needs more enhancement. Another issue we found is that the users hands and fingers were always overlaid by the virtual object which results in a confusion. Therefore, one of the solutions is to draw and render a virtual hand to overlay the virtual object. Another solution is the usage of a depth camera and applying a 3D rendering system to let the hand overlay the object when necessary. Also, in rotation, at certain angles, the index finger becomes hidden by the hand.

To evaluate that system, we conducted a set of user studies. The results were according to the performance time calculations for both experiments as well as the questionnaire’s answered by the participants. The results showed that this approach needs more enhancements regarding the performance and accuracy of gestures, yet it proved to be fun and engaging.

Future Work

As per the future work, we are looking to implement a marker less approach which needs a robust finger detection. Moreover, we need to eliminate the limitation of the positioning of the mobile device relative to the paper marker. Finally, enhancing the accuracy of the gestures; translations, rotations and scaling and support all the possible directions.

Acknowledgment

“This paper is published with partial support by the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua)”

Bibliography

[Bai et al, 2012] Bai H, Lee GA, Billinghurst M. Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, ACM, 2012, pp. 126-131

[Hürst & Van Wezel, 2013] Hürst, W. and Van Wezel, C. Gesture-based interaction via finger tracking for mobile augmented reality. Multimedia Tools and Applications, Vol. 62, Issue 1, 2013, pp.233-258.

[Seo et al, 2008] Seo BK, Choi J, Han JH, Park H, Park JI. One-handed interaction with augmented virtual objects on mobile devices. In Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, ACM, 2008, (p. 8).

[Choi et al, 2011] Choi J, Park H, Park J, Park JI. Bare-hand-based augmented reality interface on mobile phone. In Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium, 2011, pp. 275-276.

[Lee & Hollerer, 2007] Lee T. and Hollerer T., Handy AR: markerless inspection of augmented reality objects using fingertip tracking. In Proceedings of 11th IEEE International Symposium on Wearable Computers, 2007, pp. 83-90.

[Gao, 2013] Gao L. Natural gesture based interaction for handheld augmented reality. 2013.

[Baldauf et al, 2011] Baldauf M, Zambanini S, Fröhlich P, Reichl P. Markerless visual fingertip detection for natural mobile device interaction. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, ACM, 2011, pp. 539-544.

[Chun & Hollerer, 2013] Chun WH, Höllerer T. Real-time hand interaction for augmented reality on mobile phones. In Proceedings of the 2013 international conference on Intelligent user interfaces, ACM, 2013, pp. 307-314.

[Syahputra et al, 2017] Syahputra MF, Siregar RK, Rahmat RF. Finger recognition as interaction media in Augmented Reality for historical buildings in Matsum and Kesawan regions of Medan City. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Springer, 2017, pp. 243-250.

[Syahputra et al, 2018] Syahputra MF, Fatimah S, Rahmat RF. Interaction on Augmented Reality with Finger Detection and Hand Movement Recognition. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Springer, 2018, pp. 702-712.

[Ahmed et al, 2015] Ahmed L., Hamdy S., Hegazy D., El-Arif T., Interaction Techniques in Mobile Augmented Reality: State-of-the-art. In IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), 2015, pp. 424-433.

Authors' Information

Loubna Ahmed *– Ain Shams University, Faculty of Computer and Information Sciences, Computer Science Department; e-mail: lobna.ahmed@cis.asu.edu.eg*

Dr. Doaa Hegazy – *Ain Shams University, Faculty of Computer and Information Sciences, Scientific Computing Department; e-mail: doaa.hegazy@cis.asu.edu.eg*

Dr. Salma Hamdy – *Ain Shams University, Faculty of Computer and Information Sciences, Computer Science Department; e-mail: s.hamdy@cis.asu.edu.eg*

Prof. Dr. Taha I. Elarif *- Ain Shams University, Faculty of Computer and Information Sciences, Computer Science Department; e-mail: taha_elarif@yahoo.com, taha_elarif@cis.asu.edu.eg.*
Major Fields of Scientific Research: Computer Graphics, Image Processing and Artificial Intelligent.