The 5th International Conference on Electrical Engineering – Boumerdes (ICEE-B)
October 29-31, 2017, Boumerdes, Algeria.

A 3D Interaction Technique for Selection and Manipulation Distant Objects in Augmented Reality

Abdelkader Bellarbi[1], Nadia Zenati[1], Samir Otmane[2], Hayet Belghit[1], Samir Benbelkacem[1], Assia Messaci[1] and Mahfoud Hamidia[1]
[1]CDTA, Algiers, Algeria
[2]IBISC Lab, Evry University.
[1]{abellarbi, nzenati, hbelghit, sbenbelkacem, amessaci, mhamidia}@cdta.dz
[2]samir.otmane@ibisc.univ-evry.fr

Abstract—**3D object selection and manipulation is one of the essential features for any augmented reality (AR) system. However, distant object selection and manipulation still suffer from lack of accuracy and precision. This paper introduces an alternate 3D interaction technique for selection and manipulation distant 3D object in in immersive video see-through AR. The proposed interaction technique offers a high precision when selecting and manipulation distant objects thanks to the zooming-based idea. This later, allows bringing closer both real and virtual objects during maintaining the spatio-temporal registration between the virtual and the real scenes. The evaluation of our proposed approach and the comparison with other well-known techniques are given at the end of this paper.**

Keywords: Augmented reality; 3D interaction techniques; Pose Estimation; Computer vision.

I. INTRODUCTION

The increasing performance of computational and graphics hardware on both headsets and mobile devices makes the AR omnipresent [1, 2], and its applications became widespread [3, 4, 5, 6]. Thus, the interaction with the virtual objects became required. Several 3D interaction techniques in augmented reality environments have been proposed. One of their main issues is how to provide selection and manipulation of distant 3D virtual objects with precision. In addition, how to ensure the spatio-temporal registration between the virtual and the real worlds during the interaction task. However, most of those techniques still have limitations in providing intuitive and precise interaction in mobile AR.

In this work, we introduce a novel 3D interaction technique, called Zoom-In, whereby users can select and manipulate distant virtual objet in precision. This paper is organized as follows: Section 2 presents a brief related work of selection/manipulation techniques. Section 3 describes the proposed Zoom-in interaction technique. Section 4 presents results and evaluates the proposed technique. Finally, Section 5 concludes and summarizes this paper.

978-1-5386-0686-5/17/$31.00 ©2017 IEEE

II. REALATED WORKS

Several 3D interaction techniques are proposed whose goal is to provide an intuitive and a natural interaction with virtual objects. However, most of these technique are applied in virtual environment. On the other hand, selection and manipulation of distant 3D objects augmented reality are still the focus of research and interest of many researchers. Thus, most of the proposed techniques are inspired by those used for virtual reality [7]. Some of these techniques include Virtual Hand, Ray Casting, Go-Go, WIM, Image Plan, Voodoo Dolls and HOMER.

The Voodoo Dolls technique [8] allows users to scale the virtual objects by selecting a voodoo doll that has a size relative to the desired environment size. In addition, this technique allows the user to interact with distant objects by essentially bringing a representation of them closer to the user. The World in Miniature (WIM) technique [9] displays a miniature copy of the virtual world close to the user. Through direct manipulation, the user is able to interact with the miniature versions of the objects, which causes the full-sized versions of the objects to move in the same manner. However, in augmented reality, those two techniques do not ensure spatial registration of the virtual objects with the real word. Because they separate the two worlds from each other. Using Image Plan techniques [10], the user interacts with the 2D projections of the 3D objects in the scene. This allows the interaction with distant objects. Nevertheless, the interaction is limited in 2D.

The ray casting technique [11] extends a ray from the user's hand out into the desired virtual object and allows distant objects to be selected. However, it does not provide manipulations along the Z-axis, and does not allow occluded objects selection. Many techniques have been developed to improve this issue. Olwal and Feiner [12] proposed a flexible pointer that addresses the occultation problem. This technique makes it easier to point to obscured objects. The GARDEN technique [13], which based on ray-casting metaphor, allows pointing and selection virtual object in augmented reality. Yet, it does provide manipulation of distant objects. Go-Go technique [14] provides a non-linear scaling of the user's arm, which allows an interaction with distant objects. However, the non-linear amplification results an imprecise manipulation of the selected objects.

The virtual hand metaphor has been used in several works for interacting with virtual objects [15, 16, and 17]. It offers a natural interaction. However, it does not provide selection and manipulation with distant objects [18].

The HOMER technique [19] uses ray casting to select an object, once object selected, virtual hand occurs and extends to the object. The user can directly interact with this object. This provides true 6DOF manipulation of distant objects. However, HOMER cannot ensure precise interaction with small and occluded virtual objects [20]. Ha et al. [21] proposed a robust interactive augmented reality system for grabbing and manipulating 3D object. This system addresses the occlusion problem. However, it has limitations in distant object selection and manipulation.

In order to address such drawbacks, we propose in this paper, an alternate 3D interaction technique. This latter brings both of virtual and real objects closer to provide a precise selection and manipulation of 3D virtual objects. While keeping the spatial registration between the two worlds.

III. Proposed Approach

Zoom-IN is a hybrid interaction technique that combines the camera zoom for object selection, and the virtual hand metaphor for object manipulation. Thus, this technique aims to facilitate the selection and manipulation of virtual objects, in particular distant objects, while remaining linked to the real world in immersive augmented reality.

Therefore, our technique relies on the idea that the zoom of the captured images makes it possible to bring the two distant, real and virtual objects, closer together. While maintaining the spatial registration between the virtual objects and the real scene thanks to a real-time computer vision algorithm for 3D pose estimation.

A. Setup Design

In order to implement our proposed approach, we built a prototype which is composed of a "Leap Motion" controller mounted on a “Vuzix 1200AR” video-see-through headset. The HMD is equipped with an RGB camera. In order to avoid the occultation of the user's hand by the virtual objects, we rotated the Leap Motion by 45° downwards relative to the RGB camera (see Figure 1.a) [22]. Then, we aligned the virtual controller of the Leap Motion with the virtual camera. So that the user can see his virtualized hands (textured 3D hands models) on the headset display (see Figure 1.b).

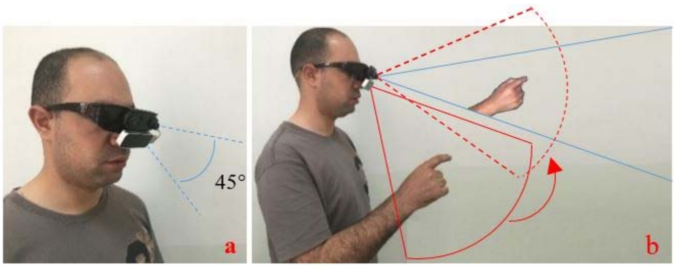

Figure 1. Zoom-In technique prototype setup. a) Leap Motion is rotated by 45° downwards relative to the RGB camera. b) 3D virtual hands are displayed on the HMD.

Therefore, in order to select and manipulate a distant 3D objet using our proposed technique. This can be performed in three steps (see Figure 2):

1. First, the user points to the desired object.
2. Then, a camera zoom is activated until the desired virtual object is close enough to be within the user's reach.
3. Thereafter, the user can use the standard Virtual Hand metaphor to grab and manipulate the object.

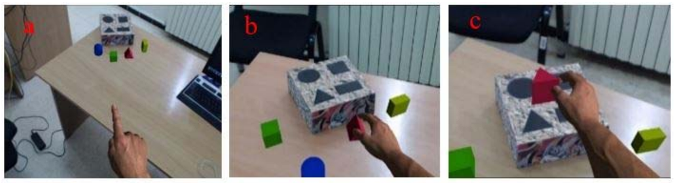

Figure 2. Zoom-In technique steps: a) Pointing to the desired object. b) Captured image zooming. c) Grabbing and manipulating the object

In order to get the distant objet at user’s hand reach, the system has to calculate first the distance between the virtual object and the virtual hand of the user. Then it estimates the zoom factor "F" to be applied to the captured image and the part of the image that must be zoomed [23]. We can formulate this problem as follow.

B. Problem Formulation

Let T_{oc} be the transformation matrix of an object O with respect to the camera C. This matrix is obtained by the product of the two matrices T_{om} and T_{cm}, which are the transformation matrices Objet-Map and Camera-Map respectively. Let Z_{oc} be the translation of the object on the Z-axis obtained from the matrix T_{oc}. (See Figure 3).

T_{hc} is the transformation matrix of the virtual hand with respect to the camera, and Z_{hc} its translation on the Z-axis extracted from the matrix T_{hc}.

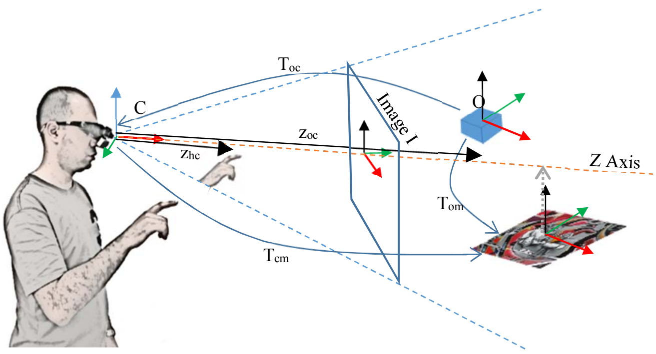

Figure 3. Zoom-IN technique architecture principle.

Let I(H, W) be the image captured by the camera with dimensions W (width) and H (height). I'(H', W') is a part of the image I such that H' / W' = H / W that represents the part of the image where the selected virtual object is projected (Figure 3).

We aim to make the selected object within the user's hand reach. This means to reduce its distance from Z_{oc} to Z_{hc}, which results by zooming the image I' with the factor $F = Z_{oc}/Z_{hc}$.

Once the zoom factor F is calculated, we can calculate the dimensions H' and W' respectively by (Equation 1 and 2):

$$H'=H/F \quad (1)$$

$$W'=W/F \quad (2)$$

On the other hand, in order to determine the part of the image to be zoomed (Figure 4), we project the 3D position O(*x, y, z*)^t of the selected virtual object on the image plane I. Then we calculate its 2D position *(u, v, 1)*^{t} with the respect to the captured image I. This results by the following equation (equation 3), which is the equation of the camera pinhole model [24].

$$s\begin{pmatrix}u\\v\\1\end{pmatrix} = A\,T_{oc}\,O\begin{pmatrix}X\\Y\\Z\end{pmatrix} \tag{3}$$

With s scale factor and A the matrix of the intrinsic parameters of the camera.

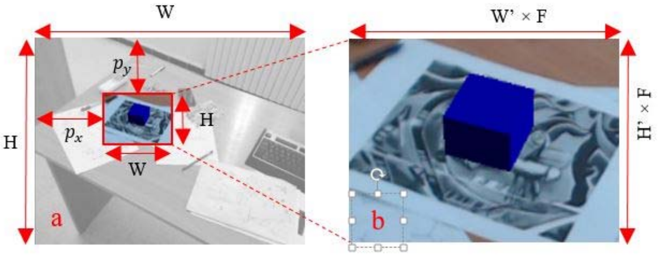

Figure 4. Extracting the portion of the image to zoom in. a) the captured image I. b) the scaled image I'.

Once we have the 2D position (u, v)t of the object on the image I, we can extract the image I' from the image I as follows (equations 4 and 5).

$$p_x = u - \frac{W'}{2} \tag{4}$$

$$p_y = v - \frac{H'}{2} \tag{5}$$

The extracted image I', can be then scaled by the zoom factor F calculated above.

C. 3D Pose Estimation

In order to ensure a stable registration of virtual objects throughout the entire augmentation process, we are based on a natural feature tracking approach. This latter uses POLAR_MOBIL detector [25] and descriptor [26] for keypoints detection and description, and CoPlanar POSIT [27] for pose estimation.

Therefore, during the interaction process. We consider the captured image I as a reference image, i.e. we change the virtual objet reference from the old map Tom to the new image reference that we call Toi. Then, for each scaled image I', we applied the natural feature tracking technique to match this image with the reference image I. The result of matching is then multiplied by Toi in order to obtain the final virtual object 3D pose. In this posture, the user can see the virtual object closer within his hand reach. Thus, he can use the Virtual Hand metaphor to grab and manipulate that object.

In addition, the proposed interaction technique allows the user to grab and manipulate neighboring virtual objects without having to repeat the entire zoom process. Furthermore, this makes it possible to select and manipulate the occluded virtual objects, which cannot be reached by other interaction technique, like Ray-casting or HOMER.

Once all the manipulations are finished, the user can terminate the interaction process by making a specific gesture by his hand (in this case, we chose the Open Palm gesture), then, a zoom-out animation is activated to restore the actual size of the captured image.

IV. Test and Evaluation

We implemented our proposed technique under Unity3D version 5.3 and Visual Studio C# 2013. We used OpenCV 2.1 and Leapmotion SDK 2.0. Running on PC with i3 3.20GHz Intel® Core™ and 6Go of RAM.

Thus, in order to test and evaluate the Zoom-In technique, we have developed a puzzle game application whose objective is to put virtual geometric shapes in their proper locations in a real box as quickly and as accurately as possible. (See Figure 5.a). We have four virtual objects with different geometric shapes: Cube, Prism, Parallelogram and Cylinder. The dimensions of the objects vary between 3 and 8 cm. The real box size is 20x20x10 cm. With four holes of different shapes on the top, which correspond to the virtual objects forms.

Figure 5. The evaluation setup. a) The box and the virtual forms used for the puzzle game. b) A participant during the evaluation process.

The main purpose of this evaluation is to compare the accuracy and the time to accomplish a task with a similar selection-manipulation technique, namely HOMER technique, which we have implemented under Unity3D with the same conditions as those of Zoom-IN.

For this purpose, we have defined the following two hypotheses:

- H0: The task completion time of the Zoom-In technique, is less than or equal to that of the HOMER technique.
- H1: The precision of the Zoom-In technique when performing a given task is better than that of the HOMER technique.

We took a group of 12 participants of different ages (26-38), different gender (4 women, 8 men) and different years of experience in the field of AR, interaction 3D and video games. All of these participants were right-handed and none of them had an identified eye problem. After explaining the principle of the game, as well as the operation of the two Zoom-In and HOMER interaction techniques, the 12 participants made familiarization tests of 5 minutes with each technique using the developed game.

Thereafter, the participants began their evaluation experiments in turns, using the developed puzzle game separately (Figure 5.b). Each participant put on the HMD (equipped by the Leapmotion) and seated at a distance of 2m from the puzzle game (the real box with the four virtual forms). Each one must repeat the experiment twice, with each interaction technique. In order to avoid a transfer of knowledge, the participants did not carry out the experiments using the two techniques in the same order.

In each experiment, we calculated the task completion time (selection time + manipulation time) for each single object. As well as the accuracy (the error in millimeters when putting the geometric shape in its location).

A. *Objective Evaluation*

We performed an ANOVA on the collected data to study the effect of the used interaction technique on time and accuracy. The results obtained were ($F = 5.321$, $p < 0.013$) for task completion time and ($f = 4.83$, $p < 0.021$) for accuracy. This reveals a significant effect of the interaction technique on these two indicators (time and precision). Overall, most of participant carried out the tasks faster and more precisely with the Zoom-In technique than with the HOMER technique (Figure 6).

Figure 6. Top: mean completion time per participant. Bottom: Mean error per participant.

B. *Subjective Evaluation*

Once all the experiments were completed, a questionnaire was given to the participants to classify the two techniques. For this end, we used the USE (Usefulness, Satisfaction and Ease) questionnaire [28].

According to the filled questionnaires, most participants appreciated the simplicity and the ease of use of the Zoom-In technique (see Figure 7). For the usefulness questions, we noticed that participants' responses were globally identical with a slight overtaking of the Zoom-In technique. Furthermore, they mentioned the ease of learning of the Zoom-In technique.

In addition, we asked the participants to give their feedback relative to the physical effort and mental stress. Most of them found that the Zoom-In technique is less tiring compared to HOMER. On the other hand, the participants categorized the two techniques nearly in the same level for the cognitive load question.

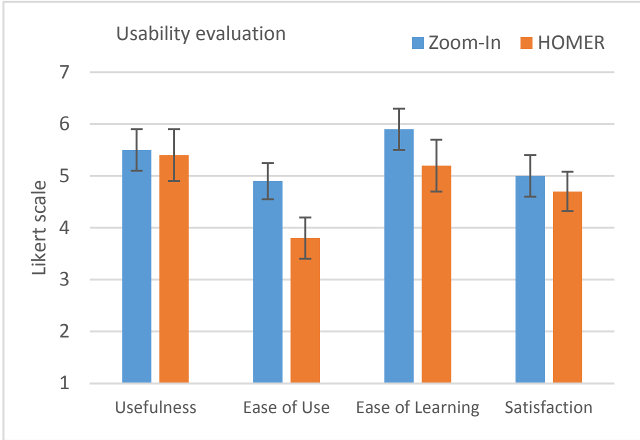

Figure 7. Usability evaluation for both of Zoom-In and HOMER techniques.

V. CONCLUSION

We have presented in this paper a novel 3D interaction technique called Zoom-IN, for selection and manipulation of virtual object in video see-through augmented reality. This technique is hybrid technique of camera zoom and virtual hand metaphor. It relies on the idea that zooming in the captured images allows bringing both of virtual and real objects closer. This technique makes it easier to select and manipulate distant virtual objects, while maintaining the spatial registration between the virtual and the real scenes, thanks to a proposed real-time pose estimation approach.

The evaluation experiments carried out on this technique have led to satisfactory and competitive results compared to HOMER technique in both precision and completion time.

REFERENCES

[1] M. Billinghurst, A. Clark, et G. Lee, « A Survey of Augmented Reality », Found. Trends Human-Computer Interact., vol. 8, no 2□3, p. 73□272, 2015.

[2] Belghit, H. Bellarbi, A. Zenati, N. Benbelkacem, S. and Otmane, S. "Vision-based collaborative & mobile augmented reality." In Proceedings of the ACM Virtual Reality International Conference VRIC 2015, p. 23.

[3] Bellarbi, A., Domingues, C., Otmane, S., Benbelkacem, S. and Dinis, A., 2012, December. "Underwater augmented reality game using the DOLPHYN". In *Proceedings of the 18th ACM symposium on Virtual reality software and technology* (pp. 187-188). ACM.

[4] Benbelkacem, S., Zenati-Henda, N., Zerarga, F., Bellarbi, A., Belhocine, M., Malek, S. and Tadjine, M., 2011. "Augmented reality platform for collaborative E-maintenance systems". In Augmented Reality-Some Emerging Application Areas. InTech.

[5] Zenati, N., Hamidia, M., Bellarbi, A. and Benbelkacem, S., 2015, March. "E-maintenance for photovoltaic power system in Algeria". In Industrial Technology (ICIT), 2015 IEEE International Conference on (pp. 2594-2599). IEEE.

[6] Benbelkacem, S., Zenati-Henda, N., Belghit, H., Bellarbi, A. and Otmane, S., 2015, May. "Extended web services for remote collaborative manipulation in distributed augmented reality". In Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on (pp. 1-5). IEEE.

[7] Krichenbauer, M., Yamamoto, G., Taketomi, T., Sandor, C. and Kato, H., 2017. "Augmented Reality vs Virtual Reality for 3D Object Manipulation". *IEEE Transactions on Visualization and Computer Graphics*.

[8] Pierce JS, Stearns BC, Pausch R. “Voodoo dolls: seamless interaction at multiple scales in virtual environments”. In: Proceedings of the 1999 symposium on Interactive 3D graphics - SI3D ’99 . New York, USA: ACM Press; 1999 p. 141–5.

[9] Stoakley R, Conway MJ, Pausch R. “Virtual reality on a WIM”. Proc SIGCHI Conf Hum factors Comput Syst - CHI ’95. 1995;265–72.

[10] Pierce JS, Forsberg AS, Conway MJ, Hong S, Zeleznik RC, Mine MR. “Image plane interaction techniques in 3D immersive environments”. ACM i3D. 1997

[11] Mine MR, Brooks FP, “Sequin CH. Moving objects in space”. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’97. New York, USA: ACM Press; 1997. p. 19–26.

[12] Olwal A, Feiner S. “The flexible pointer: An interaction technique for selection in augmented and virtual reality”. In: Uist’03 . 2003 p. 81–2.

[13] Oda O, Feiner S. “3D Referencing Techniques for Physical Objects in Shared Augmented Reality”. ISMAR 2012 - 11th IEEE Int Symp Mix Augment Real 2012, Sci Technol Pap . 2012 Nov, 207–15.

[14] Poupyrev I, Billinghurst M. “The go-go interaction technique: non-linear mapping for direct manipulation in VR”. 9th Annu ACM. 1996;79–80.

[15] Bellarbi, A., Benbelkacem, S., Zenati-Henda, N. and Belhocine, M., 2011, September. “Hand gesture interaction using color-based method for tabletop interfaces”. In Intelligent Signal Processing (WISP), 2011 IEEE 7th International Symposium on (pp. 1-6). IEEE.

[16] Bellarbi, A., Belghit, H., Benbelkacem, S., Zenati, N. and Belhocine, M., 2013, April. “Hand gesture recognition using contour based method for tabletop surfaces”. In Networking, Sensing and Control (ICNSC), 2013 10th IEEE International Conference on (pp. 832-836). IEEE.

[17] Santos, B.S., Cardoso, J., Ferreira, B.Q., Ferreira, C. and Dias, P., 2016. “Developing 3D Freehand Gesture-Based Interaction Methods for Virtual Walkthroughs: Using an Iterative Approach”. In Handbook of Research on Human-Computer Interfaces, Developments, and Applications (pp. 52-72). IGI Global.

[18] Bellarbi, A., Benbelkacem, S., Belghit, H., Otmane, S. and Zenati, N., 2014, October. “Design and evaluation of a low-cost 3D interaction technique for wearable and handled AR devices”. In *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on* (pp. 1-5). IEEE.

[19] Bowman DA, Hodges LF. “An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments”. Proc 1997 Symp Interact 3D Graph - SI3D ’97 . 1997;35-.

[20] Messaci, A., Zenati, N., Bellarbi, A. and Belhocine, M., 2015, December. “3d interaction techniques using gestures recognition in virtual environment”. In *Electrical Engineering (ICEE), 2015 4th International Conference on* (pp. 1-5). IEEE.

[21] Ha T, Feiner S, Woo W. “WeARHand: Head-Worn, RGB-D Camera-Based, Bare-Hand; User Interface with Visually Enhanced Depth Perception”. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2014. p. 219–28.

[22] Bellarbi, A., 2017. « Vers l’immersion mobile en réalité augmentée: une approche basée sur le suivi robuste de cibles naturelles et sur l’interaction 3D » (Doctoral dissertation, Université Paris Saclay; Université d'Evry Val d'Essonne).

[23] Bellarbi, A., Otmane, S., Zenati, N. and Belghit, H., 2017, March. “Design and Evaluation of Zoom-based 3D Interaction Technique for Augmented Reality”. In *19th ACM Virtual Reality International Conference (VRIC 2017)* (pp. elec-proc).

[24] McGarrity E, Tuceryan M. “A method for calibrating see-through head-mounted displays for AR”. In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99). 1999. p. 75–84.

[25] Bellarbi, A., Zenati, N., Otmane, S. and Belghit, H., 2017. “Learning moment-based fast local binary descriptor”. *Journal of Electronic Imaging*, *26*(2), pp.023006-023006.

[26] Bellarbi A, Otmane S, Zenati N, Benbelkacem S. MOBIL: A Moments Based Local Binary Descriptor. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2014. p. 251–252.

[27] Oberkampf D, DeMenthon DF, Davis LS. “Iterative Pose Estimation Using Coplanar Feature Points”. Comput Vis Image Underst . 1996 May; 63(3):495–511.

[28] Lund AM. “Measuring Usability with the USE Questionnaire”. Usability User Exp 2001;8(2).