Veridical perception of 3D objects in a dynamic stereoscopic augmented reality system

Manuela Chessa, Matteo Garibotti, Andrea Canessa, Agostino Gibaldi,
Silvio P. Sabatini, and Fabio Solari

Department of informatics, bioengineering, robotics, and systems engineering
University of Genoa, Italy
manuela.chessa@unige.it
http://www.pspc.unige.it

Abstract. Augmented reality environments, where humans can interact with both virtual and real objects, are a powerful tool to achieve a natural human-computer interaction. The recent diffusion of off-the-shelf stereoscopic visualization displays and motion capture devices has paved the way for the development of effective augmented reality systems at affordable costs. However, with the conventional approaches an user freely moving in front of a 3D display could experience a misperception of the 3D position and of the shape of virtual objects. Such distortions can have serious consequences in scientific and medical fields, where a veridical perception is required, and they can cause visual fatigue in consumer and entertainment applications. In this paper, we develop an augmented reality system, based on a novel stereoscopic rendering technique, capable to correctly render 3D virtual objects to an user that changes his/her position in the real world and acts in the virtual scenario. The proposed rendering technique has been tested in a static and in a dynamic augmented reality scenario by several observers. The obtained results confirm the improvement of the developed solution with respect to the standard systems.

Keywords: stereoscopic display, virtual reality, 3D visualization, eye tracking, dynamic stereoscopic rendering, 3D position judgment, human-computer interaction

1 Introduction

In the last decade, there has been a rapidly growing interest in technologies for presenting stereo 3D imagery both for professional applications, e.g. scientific visualization, medicine and rehabilitation system [1–3], and for entertainment applications, e.g. 3D cinema and videogames [4].

With the diffusion of 3D stereo visualization techniques, researchers have investigated the benefits and the problem associated with them. Several studies devised some specific geometrical parameters of the stereo acquisition setup (both actual and virtual) in order to induce the perception of depth in a human observer [5]. In this way, we can create stereo pairs that are displayed on stereoscopic devices for human observers which do not introduce vertical disparity, and thus causing no discomfort to the users [6]. Yet, other factors, related to spatial imperfections of the stereo image pair, that yield visual discomfort have been addressed. In [7] the authors experimentally determined the level of discomfort experienced by a human observer viewing imperfect binocular image pairs, with a wide range of possible imperfections and distortions. Moreover, in the literature there are several works that describe the difficulty of perceptually rendering a large interval of 3D space without a visual stress, since the eyes of the observer have to maintain accommodation on the display screen (i.e., at a fixed distance), thus lacking the natural relationship between accommodation and vergence eye movements, and the distance of the objects [8]. The vergence-accommodation conflict is out of the scope of this paper, however for a recent review see [9].

Besides the previously cited causes of discomfort, another well-documented problem is that the 3D position of objects, thus their 3D shape, and the scene layout are often misperceived by a viewer freely positioned in front of stereoscopic displays [10]. Only few works in the literature address the problem of examining depth judgment in augmented or virtual reality environments in the peripersonal space (i.e. distances less than 1.5 m). Among them, [11] investigated depth estimation via a reaching task, but in their experiment the subjects could not freely move in front of the display. Moreover, only correcting methods useful in specific situation, e.g. see [12, 13], or complex and expensive systems [14] are proposed in the literature. Nevertheless, to the knowledge of the authors, there are no works that aim to quantitatively analyze the 3D shape distortions, their consequences on the perception of the scene layout, and to propose an effective and general solution for an observer that freely moves in front of a 3D monitor.

In entertainment applications such distortions can cause visual fatigue and stress. Moreover, in medical and surgery applications, or in cognitive rehabilitation systems and in applications for the study of the visuo-motor coordination, they can have serious implications. This is especially true in augmented reality (AR) applications, where the user perceives real and virtual stimuli at the same time, thus it is necessary that the rendering of the 3D information does not introduce undesired distortions.

In this paper, we propose a rendering technique, the True Dynamic 3D (TD3D), and an AR system capable to minimize the 3D shape misperception problem that arises when the viewer changes his/her position with respect to the screen, thus yielding a natural interaction with the virtual environment. The performances of the developed system are quantitatively assessed and compared to the results obtained by using a conventional system, through experimental sessions with the aim of measuring the user's perception of the 3D positions of virtual objects and the effectiveness of his/her interaction in the AR environment.

The paper is organized as follow: in Section 2 we describe the geometry of the standard stereoscopic 3D rendering technique, the misperception associated with the movements of the observer, and the developed TD3D solution; Section 3 describes the AR system we have developed; Section 4 presents two experimental studies designed to validate the proposed approach; the conclusions are discussed in Section 5.

2 The geometry of the stereoscopic 3D rendering technique

2.1 The standard approach

To create the stereo 3D stimuli for the conventional approach, we have adopted the stereo rendering, based on the method known as "parallel axis asymmetric frustum perspective projection", or off-axis technique [15, 5], the technique usually used to generate a perception of depth for a human observer. In the off-axis technique, the stereo images are obtained by projecting the objects in the scene onto the display plane for each camera; such projection plane has the same position and orientation for both camera projections (see Fig. 1).

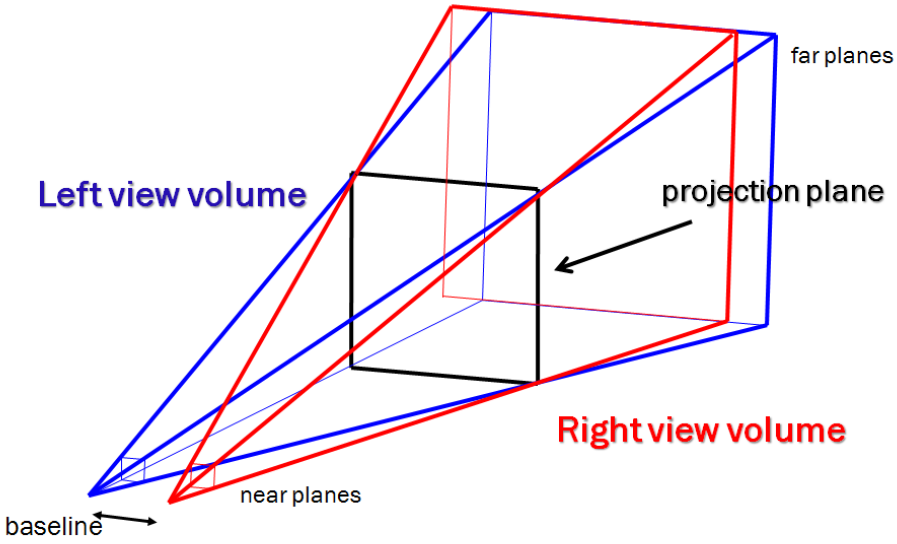


Fig. 1. The two skewed frustums for the off-axis technique.

We have also taken into account the geometrical parameters necessary to correctly create stereo pairs displayed on stereoscopic devices for human observer [5]. In particular:

- the image planes have to be parallel;
- the optical points should be offset relative to the center of the image;

- the distance between the two optical centers have to be equal to the interpupillary distance;
- the field of view of the cameras must be equal to the angle subtended by the display screen;
- the ratio between the focal length of the cameras and the viewing distance of the screen should be equal to the ratio between the width of the screen and of the image plane.

Moreover, as in [7], one should take into account the problem of spatial imperfection that could cause visual discomfort to the user, such as:

- crosstalk, that is a transparent overlay of the left image over the right image and vice versa;
- blur, that is different resolutions of the stereo image pair.

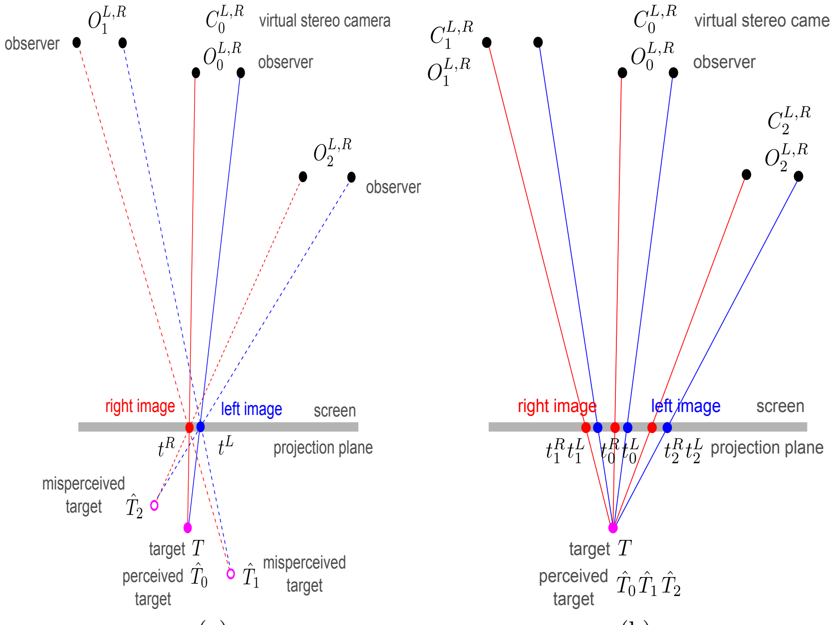
2.2 Misperception with the standard approach

The previously mentioned rules are commonly considered when designing stereoscopic virtual reality systems. Nevertheless, when looking at a virtual scene, in order to obtain a veridical perception of the 3D scene layout, the observer must be positioned in the same position of the virtual stereo camera. If the observer is in the correct position, then the retinal images originated by viewing the 3D stereo display, and the ones originated by looking at the real scene are identical [10]. If this constraint is not satisfied, a misperception of the object's position and depth occurs (see Fig. 2a).

In Figure 2a, a virtual stereo camera positioned in C_0 (for the sake of simplicity, denoting both the positions of the left (C_0^L) and right (C_0^R) cameras) determines the left and right projections t^L and t^R of the target T on the projection plane. An observer located in the same position of the virtual camera ($O_0 = C_0$) will perceive the target in a position $\hat{T}_0$ coincident with the true position. Otherwise, an observer located in a different position ($O_i \neq C_0$) will experience a misperception of the location of the target ($\hat{T}_i \neq T$). When the observer is in a location different from the position of the virtual camera, he misperceives both the position of the object and its spatial structure.

2.3 The proposed solution

To overcome the misperception problem, we have developed the TD3D rendering technique. Such a technique is capable of compensating for the movement of the observer's eyes by computing their positions with respect to the monitor, and consequently re-positioning the virtual stereo camera (see Fig. 2b). For each different viewpoint of the observer a corresponding stereo image pair is generated, and displayed on the screen. Thus, the observer always perceives the 3D shape of the objects coherently. Moreover, to maintain a consistent augmented reality environment it is necessary to have a virtual world that is at each moment a virtual replica of the real world. Thus, the screen and the projection plane should be always coincident.



(a) (b)

Fig. 2. A sketch of geometry of the stereoscopic augmented reality environment when using the standard stereo rendering technique (a), and when using the proposed TD3D approach (b). (a) In the virtual environment a target is positioned in T, and a stereo camera is placed in $C_0^{L,R}$, thus generating the left and right projections t^L and t^R on the projection plane. A real observer in the same position $O_0^{L,R}$ of the virtual camera will perceive the target correctly, whereas he/she will misperceive the position ($\hat{T}_1$ and $\hat{T}_2$) of the target when looking at the screen from different positions ($O_1^{L,R}$ and $O_2^{L,R}$). (b) In the proposed technique, the virtual camera is moved accordingly to the different positions of the observer's eyes. This yields different projections of the target (t_0^L, t_0^R; t_1^L, t_1^R; t_2^L, t_2^R) on the projection plane, thus allowing a coherent perception of the target for the observer.

3 The proposed augmented reality system

3.1 Software and hardware components

Considering the availability of commercial products with high performances and affordable costs, we decided to use off-the-shelf devices to design and develop our AR system. Specifically, we use the Xbox Kinect, a motion sensing input device developed by Microsoft for the Xbox 360 video game console. Based on an RGB camera and on an infrared (IR) depth sensor, this RGB-D device is capable of providing a full-body 3D motion capture. The depth sensor consists of an IR projector combined with a monochrome camera.

All the software modules are developed in C++, using Microsoft Visual Studio 10. To render the stereoscopic virtual scene in quad buffer mode we use the

Coin3D graphic toolkit[1], a high level 3D graphic toolkit for developing cross-platform real time 3D visualization. To access the data provided by Microsoft XBox Kinect, we use the open source driver, released by PrimeSense[2], the company that developed the 3D technology of the Kinect. The localization and the tracking of the head and the hand rely on the framework OpenNI[3], a set of open source Application Programming Interfaces (APIs). These APIs provide support for access to natural interaction devices, allowing the body motion tracking, hand gestures and voice recognition.

The processing of the images acquired by Kinect RGB camera is performed through the OpenCV 2.4[4] library.

Both the development and the testing phases have been conducted on a PC equipped with an Intel Core i7 processor, 12 GB of RAM, a Nvidia Quadro 2000 video card enabled to 3D Vision Pro with 1 GB of RAM, and a Acer HN274H 3D monitor 27-inch.

3.2 System description

Figure 3 shows the setup scheme of our system. The XBox Kinect is located on the top of the monitor, centered on the X axis, and slightly rotated around the same axis. This configuration was chosen because it allows the Kinect to have good visibility on the user, without having the Kinect interposed between the user and the monitor.

The proposed system, during the startup phase, is responsible to recognize and track the position of the skeleton of the user. After the startup phase, each time new data is available from the Kinect, the system performs a series of steps to re-compute the representation of the 3D virtual world, and to achieve the interaction with the user. These steps can be summarized as follows:

- Detection of the positions of the user's eyes and index finger, through colored markers, in the image plane of the Kinect RGB camera. This can be achieved first by obtaining the positions of the head and of the hand from the skeleton, produced by OpenNI libraries, and then performing a segmentation and a tracking of the colored markers in the sub-images, centered in the detected positions of the head and of the hand. This image processing step is based on the color information, and is performed by using OpenCV libraries.
- Computation of the position of the eyes and of the finger in real world coordinates by combining their positions in the image plane and the corresponding depths (from the Kinect D camera). The spatial displacement between the RGB and D cameras has been taken into account.
- Positioning of the stereo camera of the virtual world in the detected position of the user's eyes, obtained from the images and depth map. In order to generate correct stereo images, two *asymmetric frustums* originating from

[1] www.coin3D.org

[2] www.primesense.com

[3] www.openni.org

[4] opencv.willowgarage.com

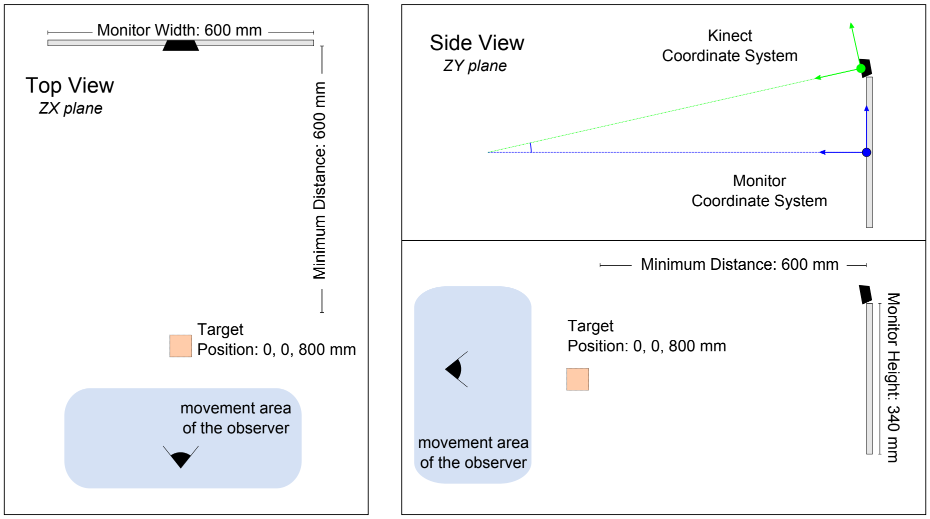


Fig. 3. The developed setup for the augmented reality system. The reported measures refer to the specific setup considered for the experimental results. The position of the target is the same of the one of the experiment shown in Figure 4.

the eyes' positions and projecting onto the same *projection plane* are created. Since the existing graphics libraries do not support the generation of such stereoscopic frustums, we have extended the behavior of the Coin3D camera class.

3.3 System calibration

To align the coordinate systems of the Kinect and of the monitor, we performed a calibration procedure by taking into consideration a set of world points, whose coordinates are known with respect to the monitor coordinate system, and their positions acquired by the Kinect. In this way, it is possible to obtain the roto-translation matrix between the Kinect coordinate system and the monitor reference frame.

The Kinect accuracy in estimating world points has been gathered from the literature [16]. In order to evaluate the precision of the algorithm, described in Section 3.2, in the estimation of the 3D position of the observer's eyes and of his/her finger, we performed a test session, by placing the colored marker at different distances, ranging from 500 mm to 1600 mm, from the Kinect sensor. For each position we acquired and processed several samples, Table 1 shows the uncertainty range on the localization of the 3D points.

Table 1. The uncertainty range (i.e. the standard deviations of the measures, expressed in mm) of the position estimation of the user's eyes and finger through our procedure that combines color based segmentation and depth map from the Kinect.

	std min [mm]	std max [mm]
X	0.22	1.16
Y	0.24	1.44
Z	0.11	1.73

4 Experiments and results

To quantitatively assess the proposed TD3D rendering technique and the augmented reality system, and to verify if it effectively allows a veridical perception of the 3D shape and a better interaction with the user, we have performed two types of experiments. In the first one, the observer was asked to reach a virtual target (i.e. the nearest right bottom vertex of a frontal cube whose width is 25 mm), in the second one, the task was to track the right bottom vertex of a cube along an elliptical path. In both cases, the scene has been observed by different positions, and we have acquired the position of the observer's eyes and of the finger during the execution of the tasks. The experiments have been performed both by using a standard rendering technique and the proposed TD3D rendering technique that actively modifies the rendered images with respect to the position of the observer's eyes. The virtual scenes have been designed in order to minimize the effects of the depth cues other than the stereoscopic cue, such as shadows and perspective. It is worth noting that since we use a real finger to estimate the perceived depth of a virtual object, we obtain a veridical judgment of the depth. If the task is performed by using a virtual finger (a virtual replica of the finger), we obtain a relative estimation of the depth, only.

For the first experiment 12 subjects were chosen, with ages ranging from 22 to 44. The second experiment has been performed by 6 subjects, with ages ranging from 28 to 44. All participants (males and females) had normal or corrected-to-normal vision. Each subject performed the two tasks looking at the scene from different positions both with the standard and the TD3D rendering techniques.

4.1 First experiment: 3D position estimation of a static object

Figure 4 shows the results of the first experiment. With the TD3D rendering technique, the positions of the observer's eyes (represented by yellow circles) are spread in a smaller range than the positions acquired with the standard rendering technique (represented by cyan circles). This happens since, as a consequence of the active modification of the point of view, the target could not be visible from lateral views. This behaviour is consistent with real world situations, in which objects that are seen through a window disappear when the observer moves

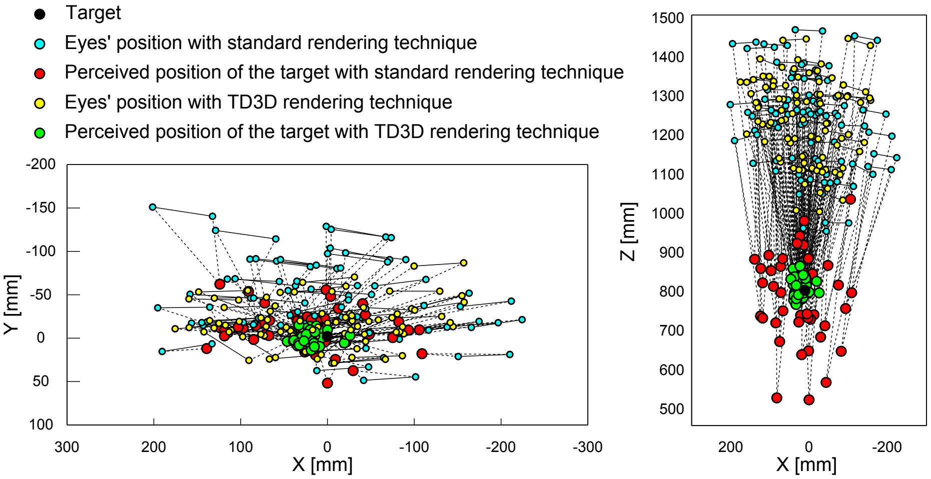


Fig. 4. Perceived positions of the target point, and the related eyes' positions for the reaching task. The perceived positions of the target with our tracking system enabled (green circles) are less scattered than the ones using a conventional system (red circles).

laterally. Nevertheless, the perceived target points with the TD3D technique and our system (green circles) are less scattered than the perceived positions of the target point with the standard rendering technique (red circles).

Table 2 shows the mean errors and their standard deviations of the perceived points. The tracking of the eyes' position of the observer and the re-computation of the stereo pairs projected onto the screen, performed by our system, yield a consistent reduction in the error of the perceived position of the target and in its standard deviation.

A graphic representation of the scattering areas of the perceived points with respect to the movements of the observer in the two situations is shown in Figure 5. It is worth noting that the area where the target is perceived, when our tracking system is enabled, is very small, thus indicating a veridical perception of the virtual object. In particular, with the standard rendering technique, the mean and the standard deviation values of the perceived 3D points, expressed in mm, are (18±65, −16±22, 775±102), whereas we obtain (21±14, −2±7, 799±23) with the developed system and the TD3D technique. Since the positions of the

Table 2. The mean error and the standard deviation values (expressed in mm) of the perceived position of the target for the first experiment. The actual position of the target is $(12.5, -12.5, 812.5)$.

	X	Y	Z
standard rendering technique	49±42	31±18	81±66
TD3D rendering technique	14±10	14±7	18±14

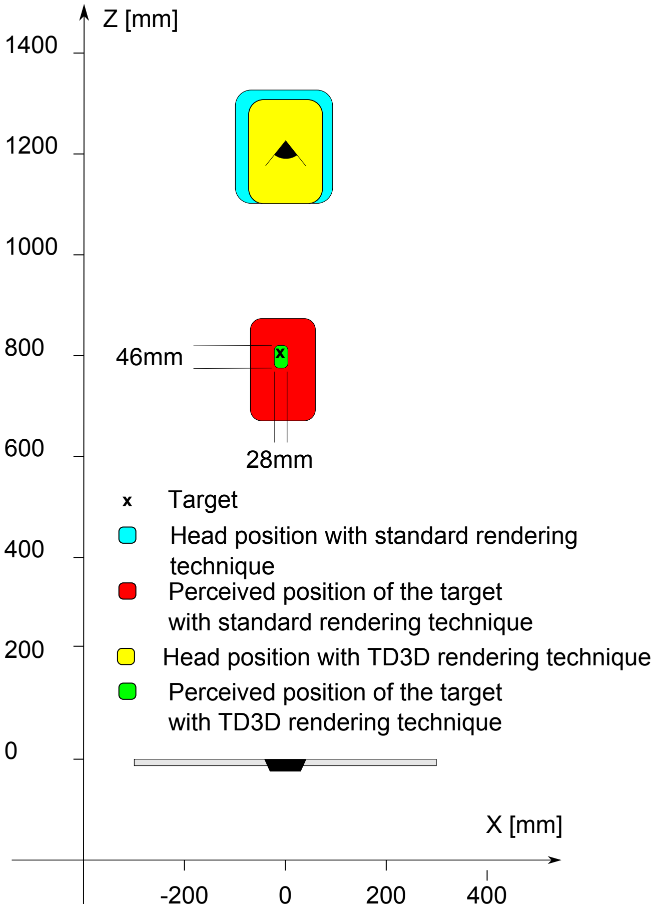


Fig. 5. Graphic representation of the results of the first experiment. The boxes are centered in the mean values of the eyes' positions and of the perceived positions of the target. The size of the boxes is represented by the standard deviation values. The smaller size of the green box represents a veridical perception (i.e. without misperception) of the position of the target.

observers are uniformly distributed in the work space, the perceived positions of the target are uniformly distributed around the actual target position (see Fig. 2), thus yielding mean values comparable between the two systems. The misperception is represented by the spread of the perceived positions of the target, that it can be quantified by the standard deviation values.

4.2 Second experiment: tracking of a moving object

In this experiment a cube whose width is 20 mm is moving along an elliptical path, from -150 mm to 150 mm in the X axis, and from 100 mm to 300 mm in the Z axis. The obtained results are shown in Figure 6. The different colors represent the acquisitions from the different points of observation. The tracks acquired with the TD3D technique are closer to the true path of the object (black line), than the tracks acquired by using the standard rendering technique. We also computed the mean errors and the standard deviations with respect to the true path of the object, obtaining a value of 50 ± 10 mm for the developed TD3D technique and of 90 ± 43 mm for the standard rendering technique.

The results of this experiment confirm that the proposed system allows a freely moving user in front of a stereoscopic display to achieve a better perception of the 3D position of virtual objects, also when they are moving.

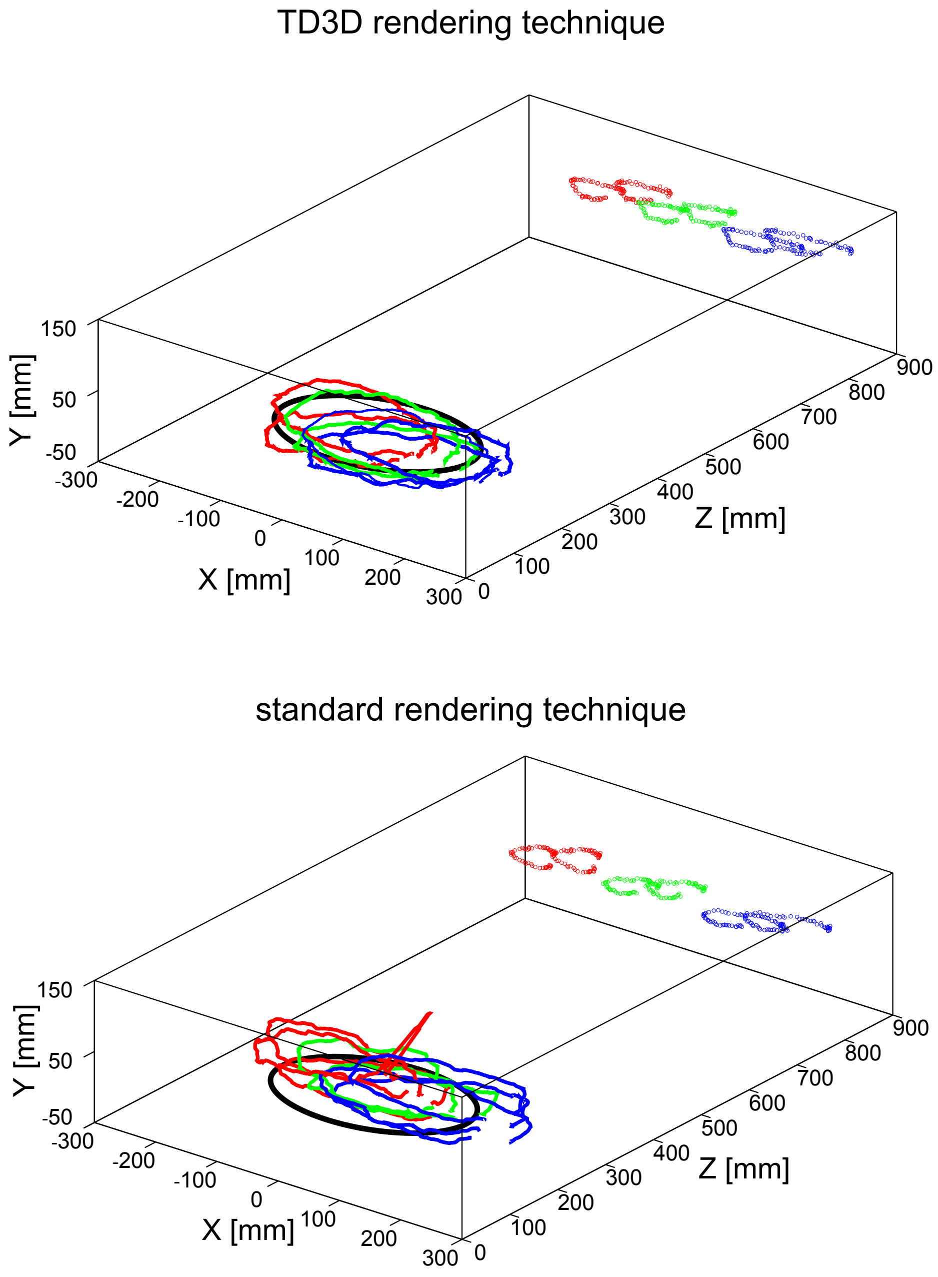
5 Conclusions and future work

A novel stereoscopic augmented reality system for natural human-computer interaction has been developed and presented. Such system allows a coherent perception of both virtual and real objects to an user acting in a virtual environment, by minimizing the misperception of the 3D position and of the 3D layout of the scene. This is achieved through a continuous tracking of the eyes' position of the observer and a consequent re-computing of the left and right image pair displayed on the screen, through a novel stereoscopic rendering technique.

In the conventional systems, when the user freely moves in front of the screen, distortions of the shape and of the distance of virtual objects occur. This issue is relevant when an accurate interaction of a real observer in a virtual world is required, especially in scientific visualization, rehabilitation systems, or in psychophysical experiments.

The proposed system relies on off-the-shelf technologies (i.e., Microsoft XBox Kinect for the tracking, and a 3D monitor with shutter glasses for the rendering) and it allows a natural interaction between the user and the virtual environment, without adding significant delay in the rendering process.

The performances of the developed augmented reality system has been assessed by a quantitative analysis in reaching and tracking tasks. The results have been compared with the ones obtained by using a conventional system that does not track the position of the eyes. The results confirmed a better perception of the 3D position of the objects obtained with the proposed system.



TD3D rendering technique

standard rendering technique

Fig. 6. Tracking of an object moving along an elliptical path, with the developed TD3D technique (top) and with the standard rendering technique (bottom). The average positions of the observers' eyes for 3 different points of observation are plotted (curves marked with circles).

Acknowledgements

This work has been partially supported by "Progetto di Ricerca di Ateneo 2010", and by the Italian National Project "PRIN 2008".

References

1. Subramanian, S., Knaut, L., Beaudoin, C., McFadyen, B., Feldman, A., Levin, M.: Virtual reality environments for post-stroke arm rehabilitation. Journal of NeuroEngineering and Rehabilitation **4**(1) (2007) 20 – 24
2. Ferre, P., Aracil, R., Sanchez-Uran, M.: Stereoscopic human interfaces. IEEE Robotics & Automation Magazine **15**(4) (2008) 50–57
3. Knaut, L.A., Subramanian, S.K., McFadyen, B.J., Bourbonnais, D., Levin, M.F.: Kinematics of pointing movements made in a virtual versus a physical 3-dimensional environment in healthy and stroke subjects. Archives of Physical Medicine and Rehabilitation **90**(5) (2009) 793 – 802
4. Kratky, A.: Re-viewing 3D implications of the latest developments in stereoscopic display technology for a new iteration of 3D interfaces in consumer devices. In: Advances in New Technologies, Interactive Interfaces, and Communicability. Volume 6616 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2011) 112–120
5. Grinberg, V., Podnar, G., Siegel, M.: Geometry of binocular imaging. In: Proc. of the IS&T/SPIE Symp. on Electronic Imaging, Stereoscopic Displays and applications. Volume 2177. (1994) 56–65
6. Southard, D.: Transformations for stereoscopic visual simulation. Computers & Graphics **16**(4) (1992) 401–410
7. Kooi, F., Toet, A.: Visual comfort of binocular and 3D displays. Displays **25**(2-3) (2004) 99–108
8. Wann, J.P., Rushton, S., Mon-Williams, M.: Natural problems for stereoscopic depth perception in virtual environments. Vision research **35**(19) (1995) 2731–2736
9. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: Predicting visual discomfort with stereo displays. Journal of Vision **11**(8) (2011) 1 – 29
10. Held, R.T., Banks, M.S.: Misperceptions in stereoscopic displays: a vision science perspective. In: Proceedings of the 5th symposium on Applied perception in graphics and visualization. APGV '08 (2008) 23–32
11. Singh, G., Swan, II, J.E., Jones, J.A., Ellis, S.R.: Depth judgment measures and occluding surfaces in near-field augmented reality. In: APGV '10, ACM (2010) 149–156
12. Lin, L., Wu, P., Huang, J., Li, J.: Precise depth perception in projective stereoscopic display. In: Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for. (2008) 831 –836
13. Vesely, M., Clemens, N., Gray, A.: Stereoscopic images based on changes in user viewpoint. US 2011/0122130 Al (2011)
14. Cruz-Neira, C., Sandin, D., DeFanti, T.: Surround-screen projection-based virtual reality: the design and implementation of the cave. In: Proceedings of the 20th annual conference on Computer graphics and interactive techniques. (1993) 135–142

15. Bourke, P., Morse, P.: Stereoscopy: Theory and practice. Workshop at 13th International Conference on Virtual Systems and Multimedia (September 2007)
16. Khoshelham, K.: Accuracy analysis of Kinect depth data. GeoInformation Science **38**(5) (2010)