Dual Autostereoscopic Display Platform for Multi-user Collaboration with Natural Interaction

Hyemi Kim, Gun A. Lee, Ungyeon Yang, Taejin Kwak, and Ki-Hong Kim

In this letter, we propose a dual autostereoscopic display platform employing a natural interaction method, which will be useful for sharing visual data with users. To provide 3D visualization of a model to users who collaborate with each other, a beamsplitter is used with a pair of autostereoscopic displays, providing a visual illusion of a floating 3D image. To interact with the virtual object, we track the user's hands with a depth camera. The gesture recognition technique we use operates without any initialization process, such as specific poses or gestures, and supports several commands to control virtual objects by gesture recognition. Experiment results show that our system performs well in visualizing 3D models in real-time and handling them under unconstrained conditions, such as complicated backgrounds or a user wearing short sleeves.

Keywords: Autostereoscopy, beamsplitter, dual display, depth information, hand detection, gesture recognition.

I. Introduction

In various fields of industry, when a new product is being designed, designers ideally should collaborate and communicate with engineers. They usually verify the design of the product (vehicles, for example) with mockups, and such verification tests are carried out repeatedly until the demands of both the designers and engineers are met. The cost and time for production increase proportionally to the number of mockup modifications. If a series of such collaborative works can be carried out in a virtual environment, using 3D displays and natural interfaces, the problems of cost and time will decrease considerably [1], [2].

Collaboration among multiple workers with virtual mockups can be held using interactive 3D displays. While there are various types of 3D displays, they usually have significant disadvantages when used in collaborative design reviewing tasks in terms of providing proximity interactions. For instance, varifocal mirrors or spinning displays usually make noise and have large space requirements, while Fog Screen and IO2 Technology's HelioDisplay have non-uniform brightness and do not provide stereoscopic images. Although autostereoscopic displays are becoming popular due to their easy access, they usually lack interactivity since they need offline processing of multi-view images.

Providing natural interaction methods for manipulating 3D virtual objects is another key aspect of a collaborative design review system. In much research, various attempts to detect hands as the first step toward recognizing gestures have been based on skin color or global image features. These algorithms can give incorrect results, such as a hand connecting with an arm if the user is wearing a short-sleeved shirt, a hand connecting with a face, or a hand connecting with another skin-colored object. Moreover, these methods produce faulty results with the presence of distractors in the background, such as humans moving around.

Recently, these problems have become solvable by utilizing depth information acquired from Microsoft's Kinect, and its tracking algorithm, which was released by OpenNI. However, extra work for initialization is needed to track the skeleton or hand motion of the user, which corresponds to the execution of a Psi pose or focus gesture, respectively.

In this letter, we propose a 3D display platform that can visualize a realistic 3D image for multiple users by adopting a

Manuscript received July 19, 2011; revised Sept. 19, 2011; accepted Oct. 5, 2011.

This work was supported by the IT R&D program of MKE/KEIT [10035223, Virtual manufacturing process verification for collaboration].

Hyemi Kim (phone: +82 42 860 1816, miya0404@etri.re.kr), Ungyeon Yang (uyyang@etri.re.kr), and Ki-Hong Kim (kimgh@etri.re.kr) are with the Creative Content Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Gun A. Lee (gun.lee@canterbury.ac.nz) is with the Human Interface Technology Laboratory, University of Canterbury, New Zealand.

Taejin Kwak (ozerodie@etri.re.kr) is with the Department of Computer Software and Engineering, University of Science and Technology, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.12.0211.0331

beamsplitter between a pair of autostereoscopic displays. Since we also provide a real-time visualization method for multi-view 3D images, users can interact with the system using gestures recognized by the proposed hand detection algorithm, which solves the problems described above.

II. Proposed 3D Interactive Display Platform

The proposed display platform is characterized by both the realistic visualization of a 3D model and natural interaction based on several gesture patterns. To make an immediate interaction possible, autostereoscopic images should be generated in real-time. Figure 1 illustrates the overall structure of the proposed system for rendering autostereoscopic images from a 3D model in real-time and providing a natural interaction method.

The proposed display platform is designed to provide a better experience for collaboration while multiple users share the display platform to view and interact with 3D virtual objects. The proposed display platform uses a beamsplitter for combining multiple autostereoscopic images displayed on a set of multi-view 3D displays using a slanted array of lenticular lenses. The display configuration provides the visual illusion of a floating 3D image. Providing not only a better environment for face-to-face collaboration, the proposed display also provides better 3D perception when used by a single user, in terms of providing a wider view volume.

As shown in Fig. 2, the proposed display platform consists of a pair of autostereoscopic displays held face-to-face horizontally with a beamsplitter placed between them [3]. The user standing in front of the platform can see the contents visualized on both displays at the same time, as the contents shown on the upper display are reflected to the user's eyes by the beamsplitter, while those on the lower display are seen directly through the mirror. As a result, the whole stereoscopic viewing area can be obtained by combining the main stereoscopic viewing area, sub-stereoscopic viewing area, and background area behind the beamsplitter. Figure 2(a) shows the front view of the display platform, and Fig. 2(b) illustrates its detailed structure. The beamsplitter has a transmission ratio of 50%. Thus, half of the incoming light is reflected, while the other half passes through, and such a characteristic provides the illusion that a multi-view stereoscopic image from the upper display is floating in air.

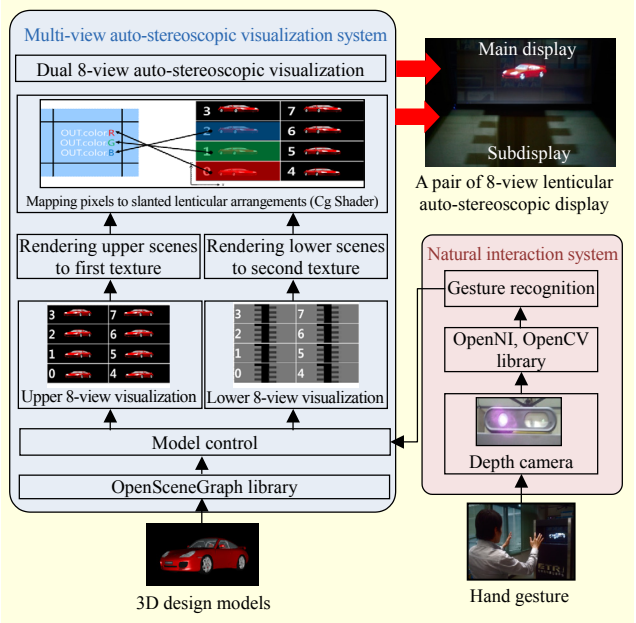

Fig. 1. System architecture of dual 3D display platform and rendering procedure.

When used for collaboration, the aforementioned display platform allows two types of data-sharing schemes. One is a widely used method in which multiple users standing side by side can see the displayed data at the same time by allotting the viewing zones (eight zones in our case) of an autostereoscopic display differently to each viewer. Taking advantage of the natural feature of an 8-view autostereoscopic display, the proposed display can provide not only different views of the same data but also different data sets, depending on the user within each viewing zone. It can be implemented by clustering adjacent views and displaying different objects to each clustered view.

Figure 3(a) shows a picture in which the display platform is used for two-user collaboration, while Fig. 3(b) shows the divided views of the display. This scheme can be applied usefully for cooperation such as sharing the same 3D structure of an object under review, in which each user can see a different representation of the same object based on their preference and needs. For instance, when two users are reviewing a 3D virtual model of a car, one can view the exterior of the car, while the other can view its interior, and still share the same spatial context with each other.

The other data sharing scheme can provide a face-to-face collaborative environment. Catching the emotional expressions from a viewer's face is one of the important decision making factors during collaborative discussions. In many cases in the

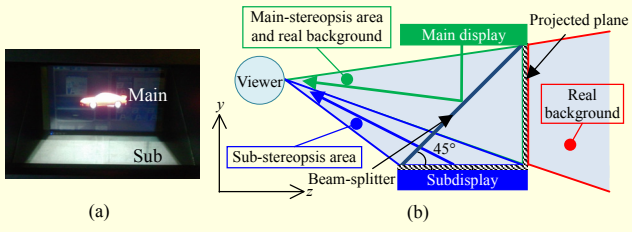

Fig. 2. Spatial expansion interface for single viewer.

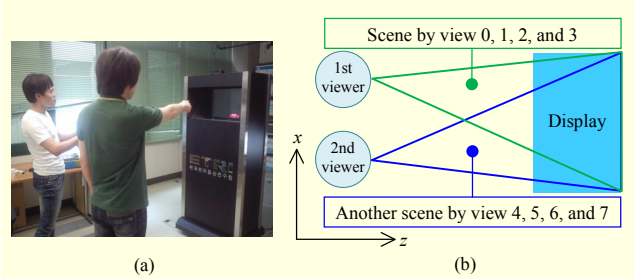

Fig. 3. Side-by-side viewing scheme for multiple viewers.

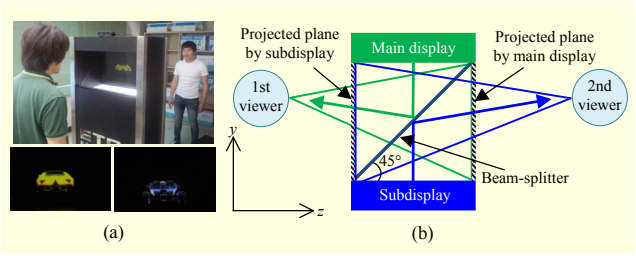

Fig. 4. Face-to-face viewing scheme for multiple viewers.

manufacturing industry, designers who have to discuss a virtual product design need to see each other's faces as well as the virtual image of the designed product. However, most modern displays, including 3D displays, have a difficult time allowing face-to-face discussions while looking at the display. In general, users usually have to look at the screen together in the same direction, which makes it hard to grasp each other's facial expressions. Under a face-to-face collaborative environment using the proposed display platform, users can see each other directly, while looking at the virtual 3D image on the display.

Figure 4(a) shows an example of how two users can share the displayed data, facing each other, while Fig. 4(b) shows how the display platform works for data viewing [3]. Under this scheme, the user in front of the display can view the front part of the object, while the other standing on the opposite side can see the back section.

The real-time autostereoscopic rendering module consists of two stages. First, it renders multiple 2D views of a 3D scene under different viewpoints into a texture. It then uses this texture with multi-view images to combine an autostereoscopic image that is supported by the display. Although the algorithm repeats the rendering multiple times (for instance, eight times in our case for a 24-inch, 8-view lenticular display), our implementation on a GPU shader, using NVIDIA Cg, provides enough real-time performance for interactivity.

To be robust to false skin-colored objects and background distractors, we utilize additional depth information as well as the color image from an RGB-D sensor released by PrimeSense. The sensor provides an image resolution of

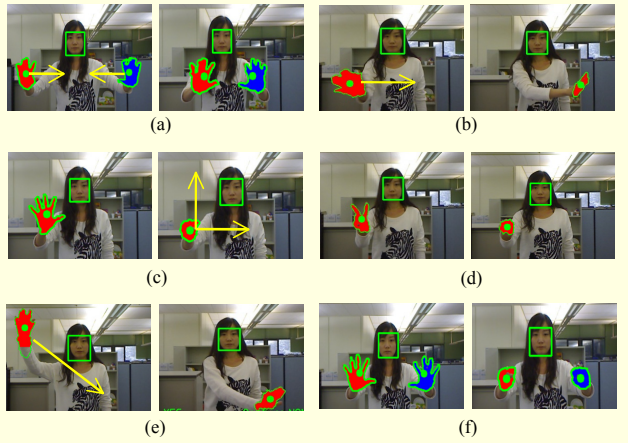

Fig. 5. Gesture commands with a left hand (blue) and a right hand (red): (a) zooming, (b) rotation, (c) translation, (d) selection, (e) undo, and (f) mode change.

640×480 pixels at a speed of 30 fps and depth information within 0.8 m to 3.5 m from the sensor.

For the first step in tracking the user's hands, we get an approximate 3D position of the user and then extract the face region using the Viola and Jones face detector. The face detector uses boosted cascades of Haar-like features. After finding the 2D position of the user's face in the color image, we obtain the position of the face in the real space using the depth map obtained from the depth camera and consider the average of the depth values as the user's depth value along the z-axis.

For hand detection, it is assumed that the hands should move within a valid region, relative to the body, where users can maximally stretch out their arms along the z-axis. The valid region for the hands is set in front of the user where the two boundaries are 20 cm and 80 cm apart from the user's body. The depth value of the user's body is based on the face detection held in the previous step. To get one or two candidate areas of the user's hands, we connect pixels using connected component labeling. To exclude the wrist or arm from the detected hand candidate area, we take the volume within a threshold distance from the closest point based on prior knowledge of the human body. Based on the depth data and prior knowledge of the human body, both hands can be detected without any devices attached to the hands or any specific initialization, and the detected hand region does not include any skin-colored objects except the hands.

To control the virtual object more freely, additional information on the fingers can help discriminate among different hand postures [4]. Finger detection is performed by evaluating the curvature of the pixels along the contour.

Figure 5 shows six intuitive hand gestures adopted for the interaction, each of which is defined as one of six commands: zoom, translation, rotation, selection, undo, and an additional

mode change. Every gesture is designed and mapped to its well-matched human action when treating real objects. While there are only six gestures defined in our current implementation, they can be expanded into a bigger set of new gestures by combining the number of open fingers and direction of the hand motion.

For collaborative scenarios, it is common to have multiple users standing in front of the interactive system. Comparing the depths of the users, we consider the user closest to the display system to be the most important, as he/she controls the virtual object. In this way, the proposed method can be considerably robust to background distractors while the main user interacts with the system through hand gestures.

III. Experiment Results

The software of the prototype system was developed with Microsoft Visual C++ 2008 and runs on Microsoft Windows 7, a 32-bit operating system. The hardware running the software is a desktop PC consisting of an Intel i7 6-core 3.33 GHz CPU, 8 GB of main memory, and a Geforce GTX 470 GPU.

To verify the real-time rendering performance of the prototype system, we carried out an experiment to investigate how the frame rate changes according to the number of polygons. The 3D data used in the experiment was of 3D models of a teapot with different numbers of polygons.

The frame rate was measured for the entire autostereoscopic rendering process, which includes rendering a 3D scene into multiple 2D images captured virtually from eight different viewpoints and composing these rendered images into a single autostereoscopic image. As shown in Fig. 6, the frame rate did not decrease until the input 3D data reached one million polygons. For 3D models with more than one million polygons, the frame rate started to decrease slowly as the number of polygons increased. However, the frame rate was still reasonable for providing real-time interactivity until the number of polygons reached two million, where the frame rate was 30 fps.

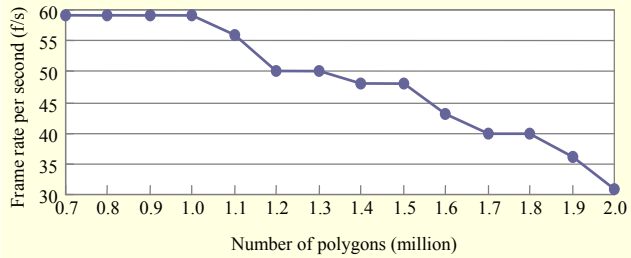

Fig. 6. Performance evaluation of frame rate corresponding to number of polygon of 3D model.

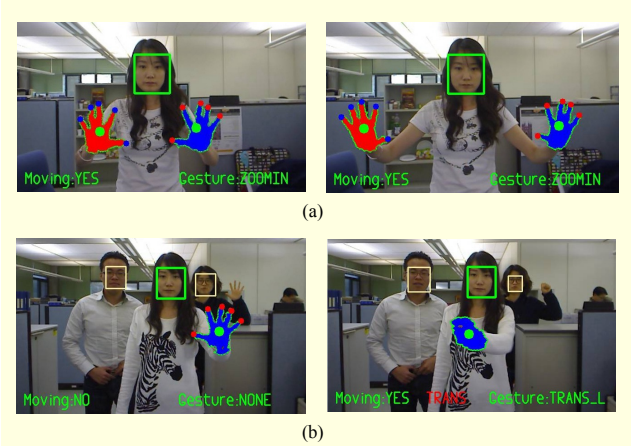

Fig. 7. Gesture recognition results: (a) wearing short-sleeved shirt and (b) existing distractors in background.

Figure 7 shows the performance of the proposed hand tracking and gesture recognition method for natural interaction in various circumstances. The users can wear short-sleeved shirts, and the background can be arbitrary (for example, no restrictions on skin-colored or moving objects).

IV. Conclusion

This letter has presented a dual autostereoscopic display platform providing a better experience of collaboration while multiple users share the display to view and interact with 3D images of virtual objects. In addition to providing a better environment for face-to-face collaboration, the proposed display platform gives an enhanced 3D perception to the user, as it can float the rendered 3D object in the air between two displays. For manipulating virtual objects naturally, the proposed system also recognizes pre-defined user gestures without any specific initialization under an uncontrolled background.

References

[1] D. Jo, U. Yang, and W. Son, "Design Evaluation System with Visualization and Interaction of Mobile Devices Based on Virtual Reality Prototypes," *ETRI J.*, vol. 30, no. 6, Dec. 2008, pp. 757-764.

[2] G.A. Lee et al., "Virtual Reality Content-Based Training for Spray Painting Tasks in the Shipbuilding Industry," *ETRI J.*, vol. 32, no. 5, Oct. 2010, pp. 695-703.

[3] U.Y. Yang, G.A. Lee, and K.H. Kim, "3D Display for Eye-Contact Collaboration," Korea Patent Applied, No. 2011-0115682, 2011.

[4] D. Lee and S. Lee, "Vision-Based Finger Action Recognition by Angle Detection and Contour Analysis," *ETRI J.*, vol. 33, no. 3, June 2011, pp. 415-422.