Real-Time Stereo Rendering Technique for Virtual Reality System Based on the Interactions with Human View and Hand Gestures

Viet Tran Hoang, Anh Nguyen Hoang, and Dongho Kim

Soongsil University
511 Sangdo-dong Dongjak-gu, South Korea
{vietcusc,anhnguyen}@magiclab.kr,
cg@su.ac.kr

Abstract. This paper proposes the methods of generating virtual reality system with stereo vision, simple and widely used 3D stereoscopic displays. However, we are motivated by not only 3D stereo display but also realistic rendered scenes popped out of screen which can be thought of as an interactive system addressing the human-to-virtual-objects manipulation. The user of the system can observe the objects in the scene in 3D stereoscopy and can manipulate directly by using hand gestures. We present the technique to render the 3D scene out of the screen and use KINECT device to keep track of user's hand movement to render the objects according to the user's view.

Keywords: virtual reality, stereoscopy, real-time rendering, head tracking.

1 Introduction

Our system presents a stereo rendering technique by which virtual models are superimposed on the computer screen and appear as realistic models in the human interactions. We present an algorithm to compute and assign the parallax value to the pair of left and right stereo images of an object in the screen space to "bring" the objects out of the screen in a whole real-time process respecting the consistence of disparity map of the original 3D rendered scene. We limit the position of the user at the distance of 40-60cm to the screen in order that the user can conveniently reach and manipulate the objects virtually.

In a stereo rendering system, it is highly required to maintain the position and orientation of the user's view toward the screen display as a straight gaze; therefore the human perception of depth and immersion can be kept stable and most accurate. Any changes in the viewer's position and direction can always potentially result in the distortion of the scene or the objects in screen space and lead the human perception into some negative symptoms of the eyes such as eye strain or fatigue. In our system, we track the human head pose, and from the information collected, we introduce a new projection matrix calculation to adjust the projection parameter for the rendering of a new scene so that the viewer cannot feel (or at least cannot perceive easily) the distortion of the scene while the 3D stereo vision can retain the fidelity.

R. Shumaker (Ed.): VAMR/HCII 2013, Part I, LNCS 8021, pp. 103–110, 2013.

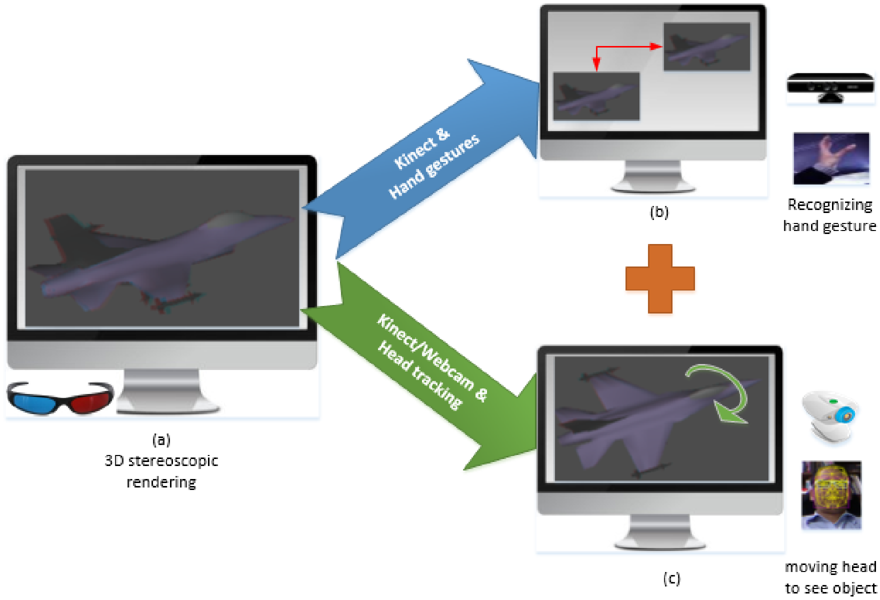

Fig. 1. Using KINECT and webcam to help the viewer to see object more realistic and interacting with them in 3d stereoscopic environment

In many cases, exposure to stereoscopic immersion techniques can be lengthy so that the user can face eye strain. The more the viewer's eyes are not oriented straight ahead to the computer screen, the more distortion of the objects will occur dramatically. We limit the angle of the user's orientation to the computer screen by 45 degree from both left and right sides from the straight direction in order to keep the rendered scene staying realistic for human perception.

In addition to enhance the interaction, we develop some basic manipulations that the viewer can perform on the virtual scene and objects. We use the KINECT device to capture the hand gestures of the viewer. We define and implement some new hand gestures based on the fundamental implementation of KINECT SDK. The viewer can perform gestures while recognizing the change of rendered scene. We define some signs of these changes in the scene objects such as they appear marked with colors when being touched; they are moving or rotating corresponding to the movement of hands.

2 Related Work

Using 3D stereoscopy in movie is a new trend in the world. There are a lot of famous movies that are produced by using 3D technology. And, some producers have spent money and time to convert their movies from 2D to 3D products. There are many researches in stereoscopic 3D to apply in many fields.

Paul Bourke wrote “Calculating stereo pairs” [10] in July, 1999 for discussing on the generation of stereo pairs used to create the perception of depth that are very useful in many fields such as scientific visualization, entertainment, games, etc. In his researches, the author was using stereo pairs with one of the major stereo 3D display technologies to create virtual three dimensional images. He already calculated eye separation distance and focal length to define category of parallax: positive parallax, negative parallax and zero parallax.

Paul Bourke created anaglyphs using OpenGL [10] in 2000 and updated these anaglyphs in 2002 by using GLUT library to do filtering automatically for left and right eye images. By using OpenGL, another person - Animesh Mishra was also rendering 3D anaglyphs [11] more precisely and effectively than previous one. He measured the amount of parallax for a vertex beyond convergence distance, calculated distance between intersections of left eye, right eye with screen for each case of parallax.

In 2008, François de Sorbier, Vincent Nozick and Venceslas Biri presented GPU-based method to create a pair of 3D stereoscopic images [7]. This is a new method using the advantages of GPU to render 3D stereo pairs including geometry shaders.

Besides, KINECT’S SDK tools support some tracking methods for using this device such as hands gesture capturing. Jens Garstka and Gabriele Peters demonstrated a view-dependent 3D projection using depth image based Head tracking method. They discussed about how to use depth image algorithm when they tracked a head. In this method, they used the depth images to find the local minima of distance and surrounding gradients to identify a blob with the size of a head, then transformed the drawn data and processed these data for tracking of a head. With view-dependent 3D projection, it provided the viewer a realistic impression of projected scene regarding to his/her position in relation to the projection plane.

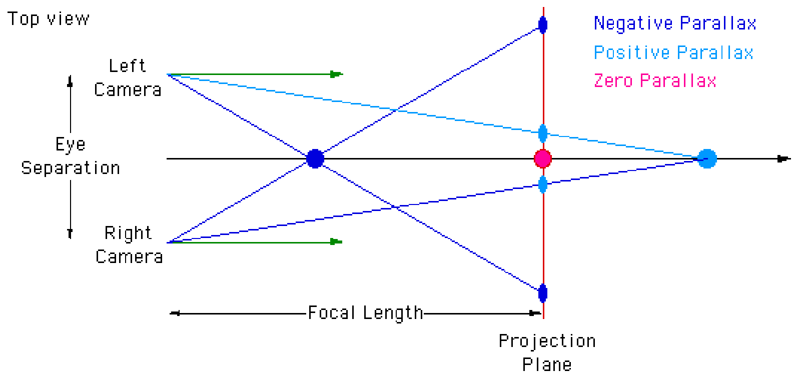

Fig. 2. Definitions in anaglyphs technology

3 Theory and Concept

3.1 3D Stereoscopy

In this paper, we focus on rendering 3D stereo using anaglyphs technology. How to render stereoscopy? In order to render stereo pairs, we need to create two images, one for each eye. We must understand some definitions about parallax, eye separation, aperture, etc.

For parallax, the distance between the left and right eye projection is called the horizontal parallax. If the object is in the opposite side from the eyes over the projection plane, it is called positive parallax. If the object is located in front of the project plane and same side with eyes, it is called negative parallax. Final definition is zero parallax where the object is located right on the projection plane.

To generate the stereoscopic images or objects on the screen, we need two images: one for left eye and one for right eye. There are two general approaches to make these images: Toe-in and Off-axis. Toe-in makes the viewer feel sick or gives some sorts of headache while Off-axis does not cause any problems. Off-axis approach is the better one and it also uses two asymmetric frustums. To get two pictures for left and right eyes, we need three steps: transforming camera (translation), calculating frustums and rendering of scene.

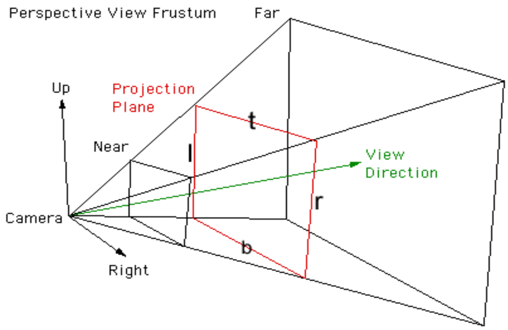

Fig. 3. Perspective view frustum

$$wd2 = near.\tan\left(\frac{\pi}{180}.\frac{FOVy}{2}\right) = t; \tag{1}$$

$$b = -\ wd2 \tag{2}$$

* Frustums for left:

$$l = b.\frac{width}{height}; \tag{3}$$

$$r = b.\frac{width}{height} + 0{,}5.\ eye\_sep.\ \frac{near}{focal} \tag{4}$$

* Frustums for right:

$$r = b.\frac{width}{height}; \tag{5}$$

$$l=b.\frac{width}{height}+0{,}5.\ eye\_sep.\ \frac{near}{focal} \quad (6)$$

As mentioned above, GPU-based geometry shaders can be used to render 3D stereoscopy by processing vertices and pixels. The main purpose of geometry shaders is to clone input primitives without requiring any process on the vertex attributes while traditional method renders it twice. By using the power of graphic cards, the performance of 3D stereo rendering in GPU-based method is approximately faster twice than the traditional methods. This method is very useful for rendering 3D stereo anaglyphs.

3.2 KINECT Tracking

There are so many researches in head pose tracking and gesture tracking. One of tracking technique is to calculate depth-images. In 2010, Microsoft launched the game controller KINECT with Xbox 360. The basic principle of KINECT'S depth calculation is based on stereo matching. It requires two images: one is captured by the infrared camera, and the other is the projected hard wired pattern [6]. These images are not equivalent because some distances between camera and projector. Therefore, we can calculate object positions in space by the view dependent 3D projection.

4 Implementation

With the expectation to reduce eye strain headache or sickness for the viewer, we implemented an application for 3D anaglyphs based on off-axis approach. Besides, to improve the performance of 3d stereoscopy, we also apply GPU-based shaders in the implementation.

Step 01: Anaglyphs using GPU

To build 3D anaglyphs, we are following the concepts mentioned in the previous part and information in "Build your own 3D display" course [3]. In this part, we define the samplers corresponding to the left and right images, then use the geometry shaders to calculate in fragment and vertex shaders of GPU and assign the output fragment color to the anaglyphs rendering in the application. Belonging to the output of left and right matrices, we have different anaglyphs mode such as full color mode, half color mode and optimized color mode. The matrix of full-color anaglyphs is shown as below:

Color Anaglyphs

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

- Partial color reproduction
- Retinal rivalry

The result of this part is an application with an object in 3D stereo belonging to anaglyphs approach. The object can be scaled, changed the position or mono mode from 3d stereo modes. (Fig.5.a)

Step 02: Tracking with KINECT

In our system, we use the 3GearSystems [9] in order to track hand gestures. This system enables the KINECT to reconstruct a finger-precise representation of hands operation. Therefore, this system gives the best results for hand-tracking process and allows us to integrate with the stereoscopic application.

3GearSystem can use either OpenNI [8] or KINECT SDK for Windows as the depth sensing camera SDK. In our system, we choose OpenNI SDK with two KINECT devices. 3GearSystem has many advantages in comparison to other systems used for hand gestures tracking. There are lots of KINECT software and algorithms working best when capturing large objects or full-body of the user. It is required that the user must stay away from KINECT sensor several meters. 3GearSystem uses two KINECTS to capture both hands and to enhance the precision of tracking process. The KINECT devices are mounted over a meter above the working place. This is to help the users work as in normal condition and they can perform hand gestures in front of their screens conveniently. This 3GearSystem satisfies our requirements.

3GearSystem uses a hand-tracking database to store the user's hand data. In order to integrate 3GearSystem into our application, we need to first calibrate the system and train 3GearSystem about our hands data. This training process consists of the following steps: calibration, training hands shape (Fig.4), training six hands poses, and creating the user data.

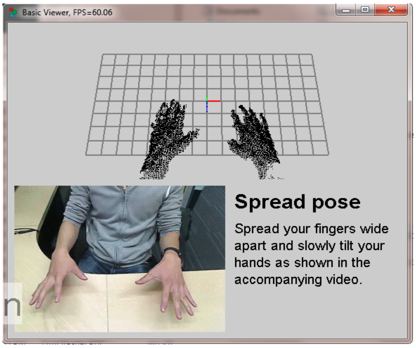

Fig. 4. Hand shape training

Step 03: Integration of the stereoscopic rendering and tracking

This paper focuses to make the viewer comfortable when they are watching the 3d stereoscopy. Furthermore, the viewer will be more interested if they interact with objects in the system or when they are moving some things by hand gestures.

We also integrated hand gestures and head tracking. With hand gestures, viewer can move an object in the system from this place to another place or from back to front of the project plane. On the other hand, head tracking will track the viewer's head when it moves left-right or up-down and displays the hidden part of an object.

This purpose will make the viewer more comfortable and reduce the eye strain or sickness. (Fig. 5(b) and Fig.5(c))

(a) (b) (c)

Fig. 5. Result of using gestures and head tracking in 3d stereo anaglyphs: (a) 3d stereoscopy; (b) moving object by using hand gestures; (c) object when head pose moves left;

We implement and run our 3D stereoscopic system on Windows 7 Ultimate version with service pack 1. Our test platform is a CPU 3.7GHz Intel Core i7, NVIDIA GTX 480 graphics card as well as 16GB of RAM and the 22-inch screen. Our system runs fast and smoothly with an average of 60 FPS.

5 Conclusion and Future Work

We present a virtual reality system with interactivity controlled by human's view and gestures using the 3D stereoscopic technology which has been popular and used a lot recently. Depending on the current technology of 3D stereoscopic display, our system can only support one viewer and has some limitation in the direction and orientation of the viewer. We demonstrate the effectiveness and the speed of our system in comparison with the related works and approaches by using GPU-based shaders for vertex and fragments. By using KINECT for tracking the gestures and head pose, this is a main idea in our work because it can make the viewer interact with system and enhance the reality when they watch the 3d object in stereoscopic via glasses. Therefore, our real-time stereo rendering system can be extended to the typical and simple virtual reality systems for learning, entertainment, etc.

In the future, we will extend our research with multiple objects displaying parallel in the system and processing hand tracking with depth images calculation. In addition, we will get the value of environment lights, then calculate the ambient, diffuse, and specular light to render anaglyph based 3d stereo objects more realistic.

Acknowledgment. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0015620).

References

1. Pollock, B., Burton, M., Kelly, J.W., Gilbert, S., Winer, E.: The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. IEEE Transactions on Visualization and Computer Graphics 18(4) (April 2012)

2. Cudalbu, C., Anastasiu, B., Grecu, H., Buzuloiu, V.: Using stereo vision for real-time head-pose and gaze estimation. U.P.B. Sci. Bull., Series C 68(2) (2006)
3. Hirsch, M., Lanman, D.: http://web.media.mit.edu/~mhirsch/byo3d/
4. Blondé, L., Doyen, D., Thierry Borel Technicolor Research, Rennes, France: 3D Stereo Rendering Challenges and Techniques. In: 44 th Conference on Information Sciences and Systems, Princeton, March, 17-19 (2010)
5. Frahm, J.-M., Koeser, K., Grest, D., Koch, R.: Markerless Augmented Reality with Light Source Estimation for Direct Illumination
6. Garstka, J., Peters, G.: View-dependent 3D Projection using depth-Image-based Head Tracking. In: 8th IEEE International Workshop on Projector–Camera Systems, Colorado Springs (June 24, 2011)
7. de Sorbier, F., Nozick, V., Biri, V.: GPU rendering for autostereoscopic displays. In: The Fourth International Symposium on 3D Data Processing, Visualization and Transmission, Atlanta, USA (June 2008)
8. OpenNI SDK, http://www.openni.org
9. 3GearSystems, http://www.threegear.com
10. Paul Bourke, http://paulbourke.net/stereographics/stereorender
11. Animesh Mishra, http://quiescentspark.blogspot.kr