An Interactive Multiview 3D Display System

Zhaoxing Zhang, Zheng Geng, Mei Zhang, Hui Dong
State Key Lab. of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

ABSTRACT

The progresses in 3D display systems and user interaction technologies will help more effective 3D visualization of 3D information. They yield a realistic representation of 3D objects and simplifies our understanding to the complexity of 3D objects and spatial relationship among them. In this paper, we describe an autostereoscopic multiview 3D display system with capability of real-time user interaction. Design principle of this autostereoscopic multiview 3D display system is presented, together with the details of its hardware/software architecture. A prototype is built and tested based upon multi-projectors and horizontal optical anisotropic display structure. Experimental results illustrate the effectiveness of this novel 3D display and user interaction system.
Keywords: autostereoscopic, multiview, interaction, 3D image, visualization, multi-projectors, anisotropic display, display system

1. INTRODUCTION

1.1 What is a "perfect" 3D display

True 3D display is the "holy grail" of visualization technology that can provide efficient tools to visualize and understand complex high dimensional data and objects. 3D Display technologies have been a hot topic of research for over a century[1][2][3][4][5][6][7][8][9][10][11][12][13][14][15][16][17][18][19][20][21][22][23].

What is a "perfect" 3D display? A perfect 3D display should function as a "window to the world" through which viewers can perceive the same 3D scene as if the 3D display screen is a transparent "window" to the real world objects. Through this window, viewer is able to perceive 3D scene with correct 3D cues, without needing to special eyewear or tracking the locations of eyes. Figure 1 illustrates the "window to the world" concept. In Figure 1a, a viewer looks at 3D objects in the world directly. We now place a 3D display screen between the viewer and the 3D scene. The 3D display device should be able to totally duplicate the entire visual sensation received by the viewer. In other words, a perfect 3D display should be able to offer all depth cues to its viewers (Figure 1b).

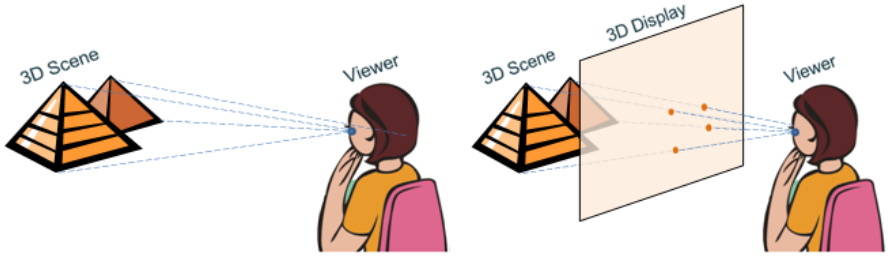

Figure 1a. A viewer looks at 3D scene directly (left) and Figure 1b. A perfect 3D display should function as a "window to the world" through which viewers can perceive the same 3D scene as if the 3D display screen is a transparent "window" to the real world.

1.2 Depth cues provided by 3D display devices

Computer graphics enhances our three-dimensional sensation in viewing 3D objects. Although an enhanced 3D image appears to have depth or volume, it is still only 2D, due to the nature of the 2D display on a flat screen. The human visual system needs both physical and psychological depth cues to recognize the third dimension. Physical depth cues can be introduced only by true 3D objects; psychological cues can be evoked by 2D images.

There are four major physical depth cues the human brain uses to gain true 3D sensation (Figure 2):

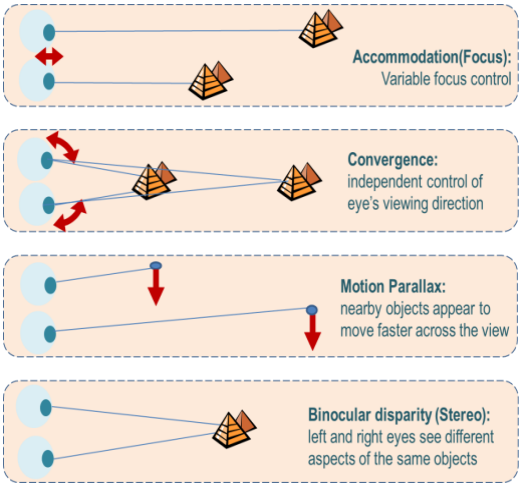

Figure 2. Illustration of four major physical depth cues

(1) Accommodation is the ability for the eyes to focus on 3D objects and to perceive the 3D depth by measuring how much the eye muscle forces eye lenses to change shape a focused image of a specific 3D object in the scene.

(2) Convergence is to measure the distance the eyes have to cross to see the 3D object in the scene simultaneously. Based on triangulation principle, the closer the object, the more the eyes must converge.

(3) Motion parallax offers depth cue by comparing the relative motion of different elements in a 3D scene. When a viewer's head moves, closer 3D objects appear to move faster than those far away from the viewer.

(4) Binocular disparity (stereo) refers to differences in images acquired by the left-eye and right-eye. The further away a 3D object, the further apart are the two images.

Some 3D display devices could provide all of these physical depth cues, while other autostereoscopic 3D display techniques may not be able to provide all of these cues. For example, the 3D Movie based on stereo eyeglasses may have difficulty in providing the accommodation or convergence, since the displayed images are on the screen, not at their physical distance in 3D space.

Human brain can also gain 3D sensation by extracting psychological depth cues from 2D monocular images[2]. Examples include (Figure 3):

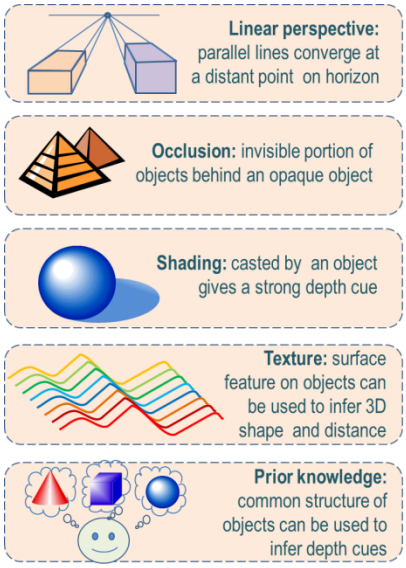

Figure 3. Illustration of psychological depth cues from 2D monocular images.

(5) Linear perspective is the appearance of relative distance among 3D objects, such as the illusion of railroad tracks converging at a distant point on the horizon.

(6) Occlusion is the invisible parts of objects behind an opaque object. Human brain interprets partially occluded objects as lying farther away that interposing ones[2].

(7) Shading casted by one object upon another gives strong 3D spatial-relationship clues. Variations in intensity help human brain to infer the surface shape and orientation of an object.

(8) Texture refers to the small-scale structures on an object's surface and can be used to infer the 3D distance of an object in the scene. The farther away the object, the coarser its texture appears to be. Parallel lines

(9) Prior knowledge of the common structures of objects – the way light interacts with their surface and how they behave when in motion – can be used to infer its depth using the other psychological cues.

Human visual system perceives a 3D scene via subconscious analysis with dynamic eye movements for sampling the various features of 3D objects. All visual cues contribute to this dynamic and adaptive visual sensing process.

It is often quite difficult for a 3D display device to provide all the physical and psychological depth cues simultaneously. Some of the volumetric 3D display techniques, for example, may not be able to provide shading or texture due to inherent transparent nature of displayed voxels. Some 3D display technologies, such as stereoscopic display, provide conflicting depth cues about the focusing distance and eye converging distance that is often referred as the accommodation/convergence breakdown.

1.3 Plenoptic Function

In 1991, Edward Adelson developed the concept of the plenoptic function[26] to describe the kinds of visual stimulation that could be perceived by vision systems. The plenoptic function is an observer-based description of light in space and time. Adelson's most general formulation of the plenoptic function P is dependent on several variables:

- the location in space from where light being viewed or analyzed, described by a three dimensional coordinate (x; y; z);
- a direction from which the light approaches this viewing location, given by two angles (θ, ϕ);
- the wavelength of the light λ, and the time of the observation t;

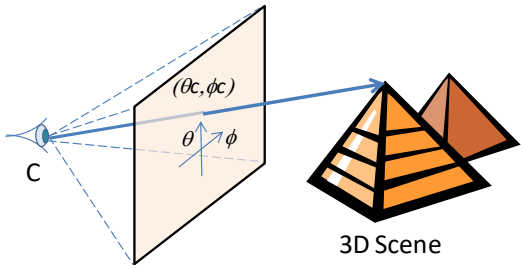

Figure 4. Plenoptic function for single viewer: The spherical coordinate system of the plenoptic function is used to describe the lines of sight between an observer and a scene.

The plenoptic function can thus be written in the following way: $P(x, y, z, \theta, \phi, \lambda, t)$. For single viewer, as shown in Figure 4, the instantaneous observation from a location C can be described in a simplified plenoptic form: $P_C(\theta_C, \phi_C, \lambda)$.

A 3D display mimics the plenoptic function of the light from a physical object (Figure 4). The accuracy to which this mimicry is carried out is a direct result of the technology behind the spatial display device. The greater the amount and accuracy of the view information presented to the viewer by the display, the more the display appears like a physical object. On the other hand, greater amounts of information also result in more complicated displays and higher data transmission and processing costs.

1.4 Directional Emitters

So far, this discussion has included only descriptions of isolated points in space emitting light of a particular color. These sources appear like stars glowing in space, never taking on a different brightness or hue no matter how or from where

they are viewed. This illumination model of omnidirectional emission is adequate to describe emissive and diffuse materials. Many natural objects are made up of materials that have an appearance that changes when they are viewed from different directions. This class of materials includes objects that have a shiny or specular surface or a microscopic structure.

Single, ideal points of such a material are called directional emitters (Figure 5): they appear to emit directionally-varying light. This definition includes points on surfaces that reflect, refract, or transmit light from other sources; the emission of these points is dependent on their surrounding environment.

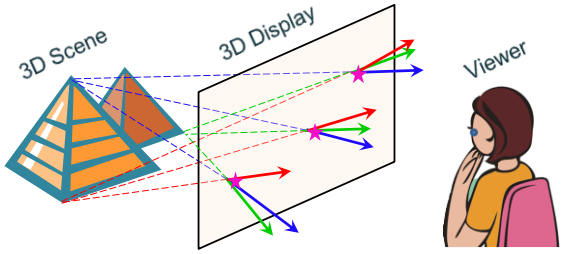

Figure 5. Each elements (voxel or hoxel) in a true 3D display should consist of multiple directional emitters: If tiny projectors radiate the captured light, the plenoptic function of the display is an approximation to that of the original scene when seen by an observer.

To model a directional emitter using an explicit depth representation requires that extra information be added to the model to describe the point's irradiance function. The light emitted by each point must be parameterized by the two additional dimensions of direction. Since a direction-by-direction parameterization would for many applications be unnecessarily detailed and expensive to compute and store, this function is usually approximated by a smaller number of parameters. The Phong light model of specular surfaces[27] is probably the most common approximation of directional emitters.

2. MULTIVIEW DISPLAY AS AN EFFICIENT 3D INFORMATION REPRESENTATION

2.1 Approximate the Light-Field by Using Multiviews

Formally, the light field[7] represents the radiance flowing through all the points in a scene in all possible directions. For a given wavelength, one can represent a static light field as a five-dimensional (5D) scalar function L(x, y, z, θ, ϕ) that gives radiance as a function of location (x, y, z) in 3D space and the direction (θ, ϕ) the light is traveling. Note that this definition is equivalent to the definition of plenoptic function. Typical discrete (i.e., those implemented in real computer systems) light-field models represent radiance as a red, green and blue triple, and consider static time-independent light-field data only, thus reducing the dimensionality of the light-field function to five dimensions and three color components. Modeling the light-field thus requires processing and storing a 5D function whose support is the set of all rays in 3D Cartesian space. However, light-field models in computer graphics usually restrict the support of the light-field function to four dimensional (4D) oriented line space. Two types of 4D light-field representations have been proposed, those based on planar parameterizations and those based on spherical, or isotropic, parameterizations.

When a viewer looks at a scene in the real world, different images are seen by left and right eyes (stereo parallax). When the viewer changes his viewing position (e.g. move his head), different sets of images are seen (motion parallax). The viewer can view a potentially infinite number of different images of the scene (continuously distributed light-field). A practical implementation strategy of light-field 3D display is to take a subsample of the continuously distributed light-field function and use a finite number of "views" to approximate the continuous light-field function.

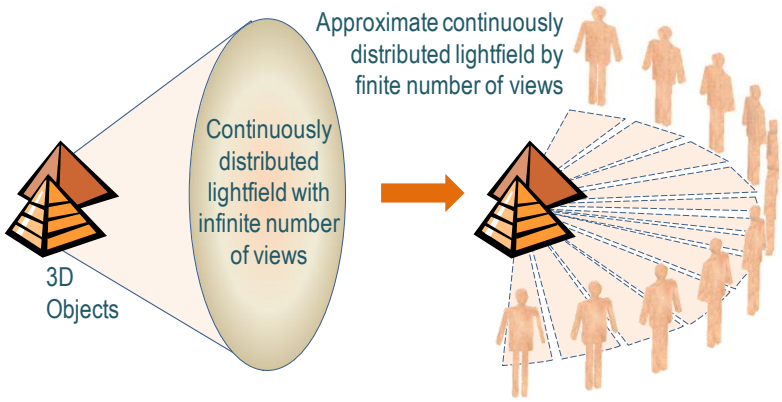

Figure 6. Use finite number of views (multiviews) to approximate the infinite number of views generated by continuously distributed lightfield.

Figure 6 illustrates the concept of using a finite number of "views" to approximate the continuously distributed light-field that in theory has infinite number of views in both directions. This approximation is viable and practical if the finite number of views is sufficiently high that exceeds the angular resolution of human's visual acuity.

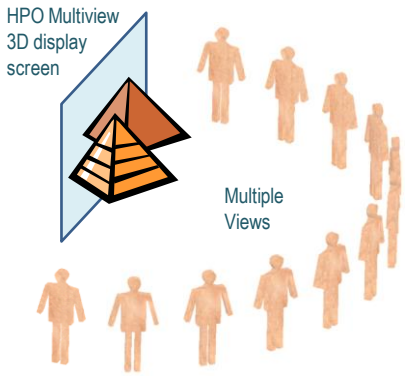

Figure 7. An illustration of multiview HPO autostereoscopic 3D display system.

Furthermore, a "horizontal parallax only (HPO)" multiview display device can be implemented that generates the parallax effect on horizontal direction only, allowing viewer's left and right eyes to see different images and different sets of images can be seen when viewer's head position moves horizontally. Even with much less number of views and with the HPO restriction, an autostereoscopic 3D display system can still generate multiviews to evoke its viewer's stereo parallax and motion parallax depth cues, thus deliver to its viewers certain level of 3D sensation (Figure 7).

2.2 Representation of 3D information for display

Depending on the type of 3D display mechanisms, there are several different types of representation methods for expressing 3D geometric model information of 3D scene. For example, in a volumetric 3D display system[14], the explicit 3D geometric model will have to be established such that each voxel is represented by its 3D coordinate values and its color components (x, y, z, r, g, b). In each refreshing cycle, the (x, y, z, r, g, b) information for all voxels need to be updated. Full geometric 3D model representation methods are usually computational expensive. For a 3D volume with 1024*768*1024 elements on each dimension, we will have a ~1 billion voxels for each frame of 3D image. Updating such a large data set in real-time (e.g. 24 frames per second) requires a data rate of ~20 GB/s. Such a data rate may present a technical challenge to existing computing hardware/software architecture.

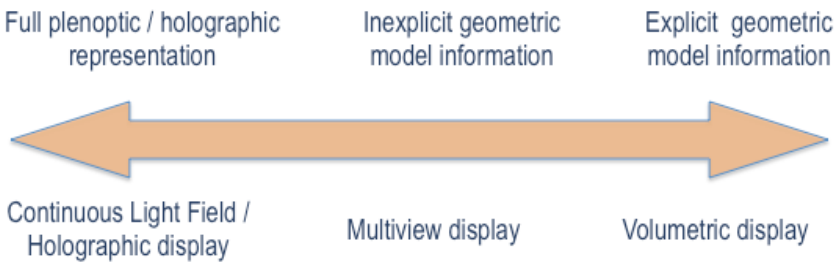

Figure 8. An illustration of various 3D model representation methods for 3D display systems.

On the other hand, light field/holographic 3D display mechanisms do not explicitly require 3D geometric model of 3D scene. Instead, these display mechanisms attempt to duplicate the entire distribution of light field or interference patterns generated by 3D objects in the scene. Highly fidelity 3D display effect can be generated, with the expensive of tremendous amount of computation effort. Engineering a practical 3D display system that is able to achieve continues light field distribution or to duplicate entire interference pattern is very challenging tasks. To date, no one has ever been able to achieve this ultimate goal.

A practical compromise to deliver sufficient 3D sensation to viewers without requiring enormous computation power can be achieved by the multiview 3D display mechanisms (Figure 8). Using discrete approximation of continuous light field distribution makes it possible to build multiview systems with relatively less expensive components. For example, the data rate for a 64-view XGA multiview system with an updating rate of 24Hz is 1.2GB/s (1024*768*64*24). This data rate is within the reach of existing hardware/software technologies.

Another advantage of multiview 3D display is that its data format is compatible with existing TV signal transmission infrastructure, which is view-based, instead of geometric model based infrastructure. In the foreseeable future, multiview 3D display will represent a promising strategy for 3D display and 3D TV applications.

2.3 Multiview image formation mechanisms

The effect of multiview display is produced by combination of two functional modular: (1) Image generation: Broadly speaking from image generation point of view, there are two basic types: Projection-based v.s. Flat-panel-based multiview display systems; (2) screen with required optical property: so that it can deliver the image for each individual view to its corresponding direction. There are a variety of screens that can facilitate the multiview 3D display. For example, lenticular lenses, parallax barriers, vertical diffusers, holographic plates, micro-lens array (integral imaging) screens.

The field of view, or "viewing zone", for multview 3D display is defined as the spatial region within which viewers are able to see complete 3D image displayed on the screen. For a multiview 3D display system, the viewing zone is divided into K "views". Within each view direction, reviewers can only see one image generated by the multiview 3D display system. Images for different views are corresponding to the different 3D perspective views of a 3D scene to be displayed on the screen. There images usually have similar sizes. The overlapping between images from neighborhood views should be minimized. In a HPO system, the horizontal separation between neighboring views should not be greater than the inter-pupillary distance of human eyes (typically 62 mm) at the viewers' location. Otherwise, the viewers cannot see different images from different views, therefore are not able to perceive the 3D depth sense based on stereo parallax (binocular parallax). When viewers change their viewing location, different sets of images are presented to their eyes from different views, thus producing the motion parallax.

In this article, we restrict our discussions mainly on the HPO type of displays. The projection-based multiview systems use multiple projectors to generate images corresponding to multiple view directions, and then project these images onto a 3D image formation screen. Figure 9 shows two simplest methods of creating horizontal parallax only (HPO) multiview 3D displays using multiple projectors. Each of these projectors creates images for a single view, respectively. In Figure 9a, the projections form images on lenticular reflective screen. In vertical direction, the projected light is diffused in all the directions. In horizontal direction, the projected light is focused onto the reflective diffuser screen and then projected back to the direction of the projector. Light passes through the lenticular screen twice. In the first pass, the lenticular focuses projected image on the diffuse screen, and in the second pass, the same lenticular redirect the image

back to the same angular direction. A transmissive screen can also be used, together with a double lenticular sheet, as shown in Figure 9b. Vertical diffuser or holographic films can also serve as the screen materials.

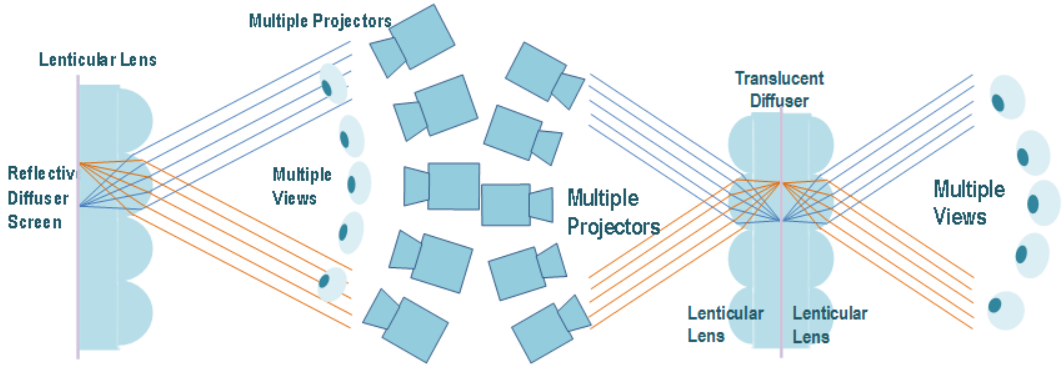

Figure 9a. A autostereoscopic 3D display using multiple projectors (frontal projection)

Figure 9b. A autostereoscopic 3D display using multiple projectors (rear projection)

On flat-panel-based multiview 3D displays, the images are produced by a single flat panel of display components, such as LCD.

This article focuses on the projection-based multiview technologies and our primary contributions that will be elaborated later are as follows:

- Present a easily- commercialized autostereoscopic multiview 3D display system that could provide large color light-field 3D images;
- The real-time interactivity gives the observers stronger 3D perception of the virtual object;
- Improved algorithm are applied to satisfy real-time application rendering;

3. INTERACTIVE MULTIVIEW 3D DISPLAY SYSTEM

3.1 System overview

1) Screen

Our system primarily consists of one server personal computer (PC), four client PCs, one 32-projector-array and one anisotropic screen. The lenticular structure (Figure 10, 11) contributes to the anisotropic effect of screen as it reflects the incoming light towards the same direction in each viewing zone.

In the front projection case that we apply, the light entering the lenticular structure focuses on the front surface of the diffused substrate, which makes the length of the back focal plane zero. As the known equation (1) shows[28]:

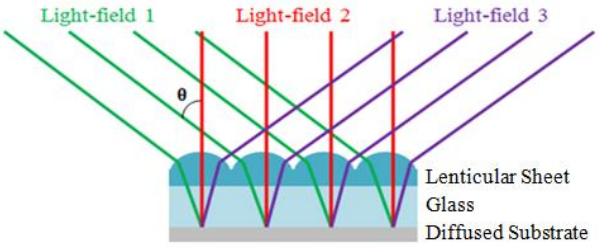

Figure 10. Illustration of front projection screen structure.

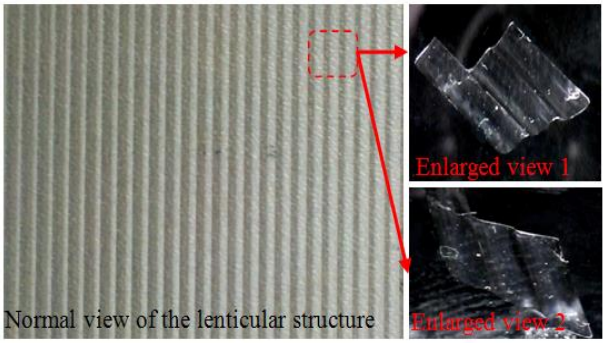

Figure 11. The front projection lenticular screen structure.

$$L_{BFD} = \frac{r}{n-1} - \frac{e}{n} \tag{1}$$

where r represents the radius of each lenticular unit, n the material reflection coefficient and e the thickness of the lenticular sheet and glass, $L_{BFD} = 0$ leads to[28]:

$$e = r\frac{n}{n-1} \tag{2}$$

Limited to the existing standard glass specifications, we actually choose 5mm instead of the precise value from equation (2) for e the thickness value.

2) Projector array

The projector-array that consists of 32 LED projectors (Figure 12) is setup according to the angle between two adjacent light-field (represented as θ in Figure 10). In most cases, projectors in compact form are preferable as θ is commonly no larger than 1 degree. Limited to the multi-output ability of the current display card, Client-Server (C/S) model is applied in this system to increase the number of the light-field image projectors. Furthermore, sharp jump of the display contents occurs as the observers move across different viewing zones. The number of the viewing zones is determined by the field of view (FOV) of the lenticular screen[8]. Identical content can be seen in each viewing zone so we keep our discussion in a single viewing zone in the following paragraphs.

3) Hardware system architecture

As Figure 13 demonstrates, the Server PC controls four client PCs (C1 to C4) to feed the projectors (P1 to P8) with light-field images respectively, enabling a total of 32-viewpoint autostereoscopic multiview 3D display. A camera is connected to the server to perform the multiview-alignment and interactivity. Ignoring the unavoidable pixel loss in the multiview-alignment procedure, the display-fresh rate of the Video stream in this system is 60Hz with 1024*768 pixels per image.

Figure 12. Projector-array

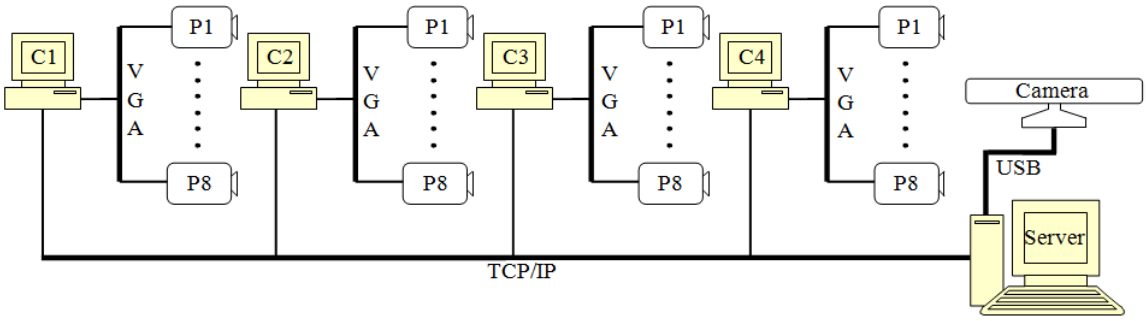

Figure 13. Hardware system architecture of the interactive multiview 3D display system.

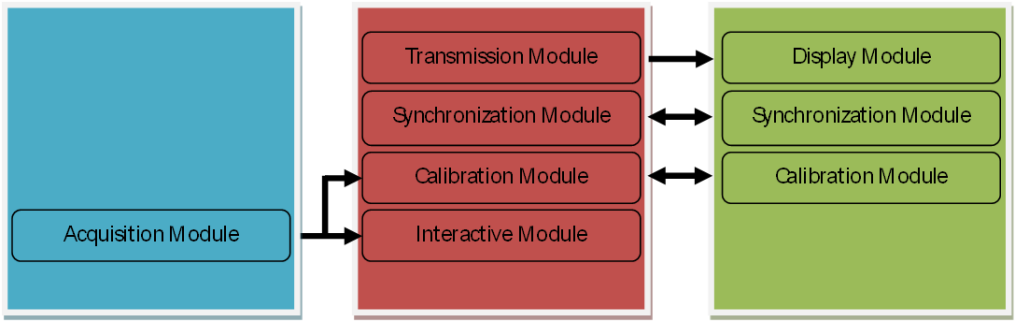

Figure 14. Modules of the software of the interactive multiview 3D display system.

4) Software architecture

The software modules in this system are shown in the Figure 14. The server PC firstly uses real-time acquired images from the camera to deal with the calibration of the light-field images. And the client PCs are able to produce smooth-transitioned light-field scenes using the calibration data. Both the calibration and the display modules in the client PCs follow the instructions from the server through the transmission module. And the interactive module is responsible for the interactivity between the light-field scenes and the manipulator.

As all the applications are based on the C/S model, the synchronization module is indispensable to make robust communication between the server PC and the client PCs. Since we create independent threads for each of the display process in the display module to reduce the rendering delay in different projectors, the synchronization module is needed to maintains that these display threads work together well with other critical threads.

3.2 Alignment of multiview projection images

As more than one projectors are involved, the distortion correction is indispensable for the projected images from all these projectors. Generally, projectors apply trapezoid light-path to gain larger FOV. In this case, projected images from projectors positioned horizontally different will form perspective-warped quadrilaterals instead of rectangles on the display plane. To make all these non-uniform images overlapping completely and fit into the screen, we firstly adjust the projectors to maximize the overlapping part of these quadrilaterals, where the screen is placed (Figure 15). Being contained completely inside in any of the four quadrilateral-shaped projected images (marked as No. 1-4) , the screen then servers as the "benchmark-rectangle" for all the perspective-warped images to transform to (the warped images represented by No. 1-4 have been transformed to the same size to match the screen rectangle).

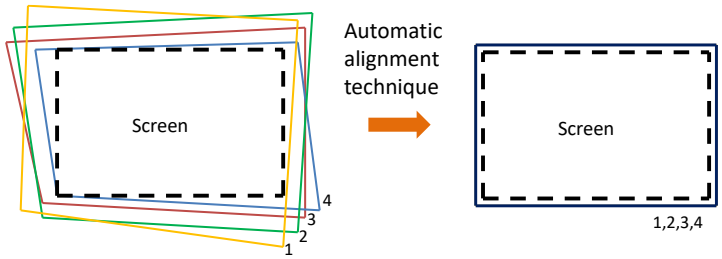

Figure 15. Screen with projected images before and after the correction using automatic alignment technique we developed.

Since the display screen is flat, the projection alignment of multiple images only needs to match four points in each images to four corners in the screen rectangle. As long as all quadrilateral-shaped images share the same "benchmark-rectangle", it is possible to align all these projected images to the screen. We thus developed the camera-based automatic alignment software to achieve this goal. Figure 16 shows the complete automatic multiview projection alignment procedure.

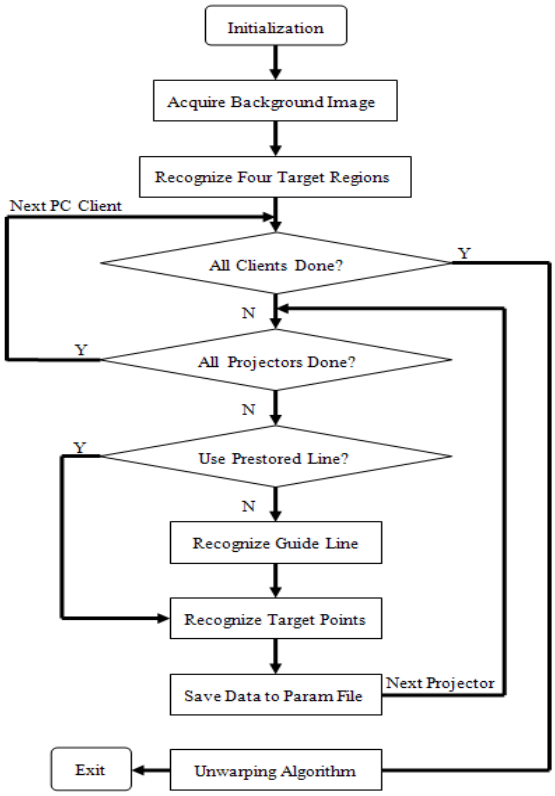

Figure 16. Flow chart of the automatic alignment in our system.

Two key parts in this automatic procedures are the acquirement of the four matching pairs of points and the transform algorithm we use to accomplish the alignment.

1) Finding four matching pairs of points

In automatic condition, the computer needs to know the coordinates of the four points drawn on the projected images in time once they are to be "seen" at the four corners of the screen. Therefore, a camera is involved to tell if the points are in their right places.

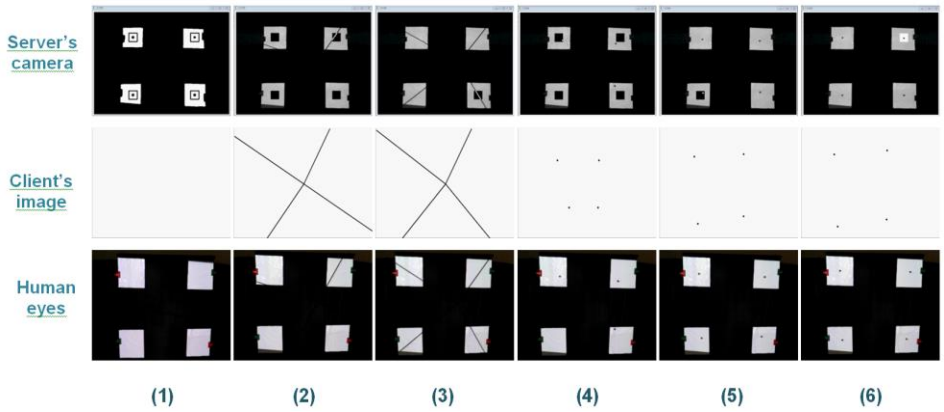

Figure 17. Procedures of finding four matching pairs of points.

We initially draw the four points in the center of the projected image and move them step by step towards four corners of the image. The anisotropic characteristic of the screen will give rise to sever interference as it reflects so much to dazzle the camera. We solve this by putting four diffused reflection surface in the corners of the screen as the regions of interest (ROIs) for the camera. Again, to make sure that the points will ultimately move into the four pre-set ROIs, we draw four guide lines from the center to each corner of the image and by moving them clockwise the camera will recognize these guide lines when they sweep through the ROIs. The points will then be moved into the ROIs along their corresponding guide lines. Then the camera only needs to focus on the four ROIs and record the coordinates of the four points in the image when they are in the corners of the screen. In a word, we achieve the transformation of the screen's four corner-points from the camera's coordinate system to the image's coordinate system. Figure 17 shows how this is performed in views of the server's camera, the client's image and the human's eyes.

(1) Recognizing four ROIs in the four screen corners, with white-color projected image as backlight;

(2) Drawing and "spinning" four guide lines;

(3) Recognizing the four guide lines in their corresponding ROIs;

(4) Moving four points towards the corners of the screen along the recognized guide lines;

(5) Recognizing all the four points once they enter their corresponding ROIs;

(6) Moving four points to the exact positions of the screen corners and record the critical data;

2) Correcting the perspective-warped images

In this stage we perform the perspective transformation with the acquired coordinates of the screen's four corners in the image's coordinate system. Four pairs of reference transformation points are thus grouped: for the image (U_i, V_i) (i = 1, 2, 3, 4) and for the screen (X_i, Y_i) (i = 1, 2, 3, 4). According to the mathematical description of the perspective transformation, we present the equation (1), with U, V representing the coordinates of transformed space, X, Y the coordinates of the original space and w the constant parameter[29]:

$$\begin{bmatrix} wU \\ wV \\ w \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \\ c_{22} \end{bmatrix} = M_{coefficient} \begin{bmatrix} X \\ Y \\ c_{22} \end{bmatrix} \tag{3}$$

We could deduce equation (3) further with the known condition that c22 = 1:

$$U_i = \frac{c_{00}X_i + c_{01}Y_i + c_{02}}{c_{20}X_i + c_{21}Y_i + 1}$$

$$V_i = \frac{c_{10}X_i + c_{11}Y_i + c_{12}}{c_{20}X_i + c_{21}Y_i + 1} \tag{4}$$

In the equations (4) above, we have eight unknown variables c_{00}, c_{01}, c_{02}, c_{10}, c_{11}, c_{12}, c_{20}, c_{21}. And with the four acquired four pairs of reference points, we could get the coefficient matrix through the equation (5):

$$\begin{bmatrix} X_0 & Y_0 & 1 & 0 & 0 & 0 & -X_0U_0 & -Y_0U_0 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -X_1U_1 & -Y_1U_1 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -X_2U_2 & -Y_2U_2 \\ X_3 & Y_3 & 1 & 0 & 0 & 0 & -X_3U_3 & -Y_3U_3 \\ 0 & 0 & 0 & X_0 & Y_0 & 1 & -X_0V_0 & -Y_0V_0 \\ 0 & 0 & 0 & X_1 & Y_1 & 1 & -X_1V_1 & -Y_1V_1 \\ 0 & 0 & 0 & X_2 & Y_2 & 1 & -X_2V_2 & -Y_2V_2 \\ 0 & 0 & 0 & X_3 & Y_3 & 1 & -Y_3V_3 & -Y_3V_3 \end{bmatrix} \begin{bmatrix} c_{00} \\ c_{01} \\ c_{02} \\ c_{10} \\ c_{11} \\ c_{12} \\ c_{20} \\ c_{21} \end{bmatrix} = \begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ U_3 \\ V_0 \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} \tag{5}$$

Once acquiring the coefficient matrix, we could perform the perspective transformation through not only the four reference point pixels but all the pixels in our target images. A complete calibration process for a 1024 by 768 image to be corrected should be as follows:

$$U_j = \frac{c_{00}X_j + c_{01}Y_j + c_{02}}{c_{20}X_j + c_{21}Y_j + 1}, 1 \le j \le 1024 \times 768$$

$$V_j = \frac{c_{10}X_j + c_{11}Y_j + c_{12}}{c_{20}X_j + c_{21}Y_j + 1}, 1 \le j \le 1024 \times 768 \tag{6}$$

The OpenCV Library provides a set of functions to perform this transformation. However, it's not efficient enough in some real-time rendering applications. Noticeably, different frames of the same projectors share the identical $M_{coefficient}$ and U_j, V_j in equations (3) and (6). We thus set up the look up table (LUT) for each projector to save the results the computer calculates for the first frame and assign the stored data to all the later frames. Equations (7) show that this replaces 5 multiplication and 4 addition operations by only one assignment operation, reducing computation time to a considerable scale.

$$U_j = \frac{c_{00}X_j + c_{01}Y_j + c_{02}}{c_{20}X_j + c_{21}Y_j + 1} = LUT_j.U, 1 \le j \le 1024 \times 768,\ frame = 1$$

$$U_j = LUT_j.U, 1 \le j \le 1024 \times 768,\ frame > 1$$

$$V_j = \frac{c_{10}X_j + c_{11}Y_j + c_{12}}{c_{20}X_j + c_{21}Y_j + 1} = LUT_j.V, 1 \le j \le 1024 \times 768,\ frame = 1$$

$$V_j = LUT_j.V, 1 \le j \le 1024 \times 768,\ frame > 1 \tag{7}$$

In addition, assembly instead of C++ programming language could be applied to deal with the traversing of the whole pixels in the image in order to boost the processing efficiency further, although this is not as effective as the usage of LUT because the OpenCV Library also ameliorates this part of its functions. The following table concludes all the computation time of the methods we have tested in our experiments:

Table 1. Comparison of computation time between different perspective transformation methods

	Custom	OpenCV Library	Custom with LUT	OpenCV Library with LUT	Custom with LUT in Assembly	OpenCV Library with LUT in Assembly
10 Frames (1024*768)	2.125s	1.457s	1.375s	0.437s	0.203s	0.189s
Average	0.213s	0.146s	0.138s	0.044s	0.020s	0.019s

4. EXPERIMENT RESULTS OF THE INTERACTIVE MULTIVIEW 3D DISPLAY SYSTEM

4.1 Static multivew 3D display

In Figure 18, we firstly present the light-field image series of the same object rendered at two different angular intervals: the 0.5-degree and 1.5-degree (Due to limitation of space, we pick images taken at 8 viewing point out of 32):

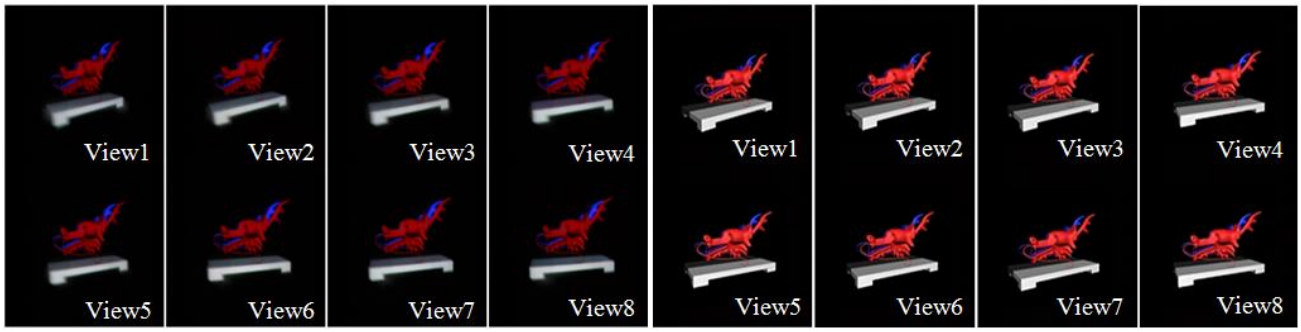

Figure 18a. 0.5-degree angular interval displayed and rendered

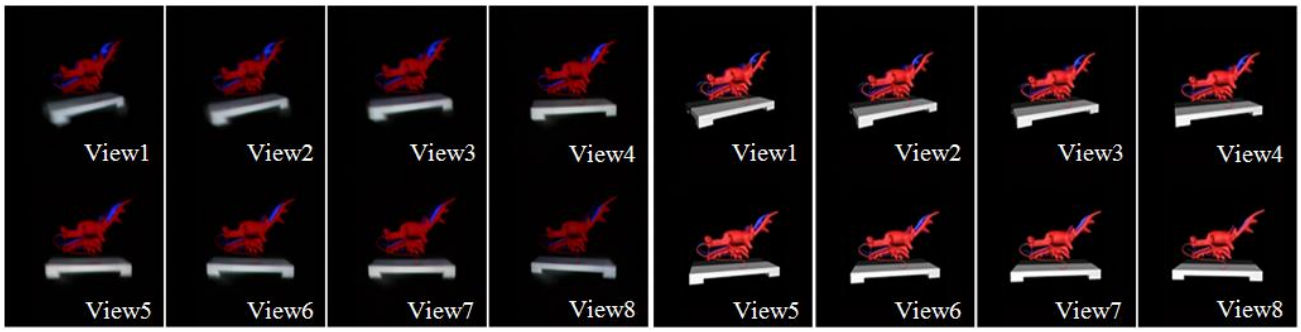

Figure 18b. 1.5-degree angular interval displayed and rendered

As the comparison of light-field images in terms of angular interval, we find that on the one hand the larger angle-interval gives rise to the more evident HPO effect as the contents in the first and the last of the image series differ more in the 1.5-degree angular interval group than those in the 0.5-degree angular interval. On the other hand, the decrease of the angular interval will proportionally restrain the crosstalk between the images viewed adjacently. Accordingly, we can gain the multivew 3D image with less ghosting with the 0.5-degree angular interval and more with the 1.5-degree.

In addition, our system is able to display the real object captured by camera-array placed the same way we set our projector-array. In Figure 19, a group of 16-position light-field test images – "the Stanford Bunny" by courtesy of Stanford Computer Graphics Laboratory is demonstrated by our system. The strong restrain on ghosting and subtle HPO effect (can be observed from the change in the bunny's right eye) due to the very small sampling angular interval matches our discussion above. (Limited by space, only images at 8 viewing points out of 16 are listed.)

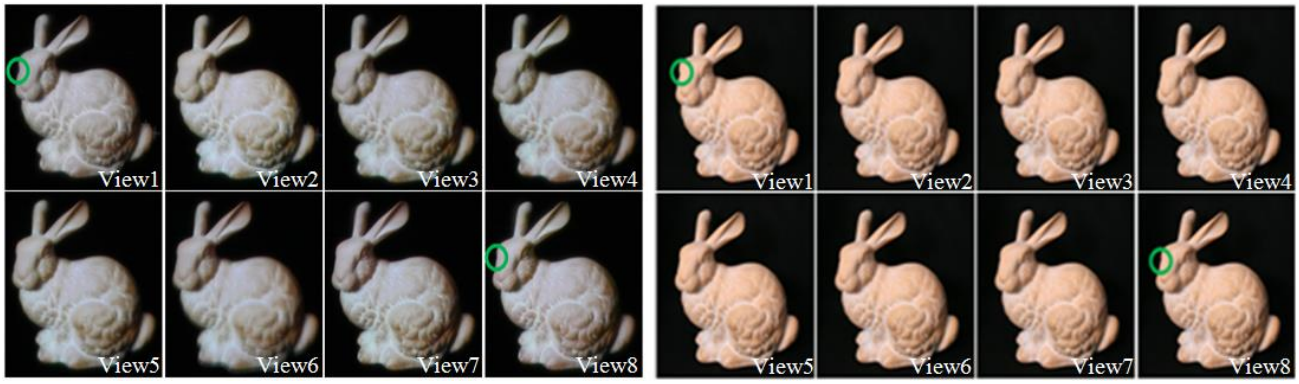

Figure 19. The Stanford Bunny displayed and captured

4.2 Interactive multivew 3D display

We render a 3-by-3 Rubik's cube and define several human gestures for the observer to manipulate. Apart from what have been summarized in Figure 14, the finite-state-machine-like (FSM-like) thought is applied to the interactive module as the server PC updates each frame of the captured human hand image, analyses the current display state and send commands to all the client PCs according to the recognized gestures from the manipulator. The complete procedure in both server PC and client PCs is shown in Figure 20:

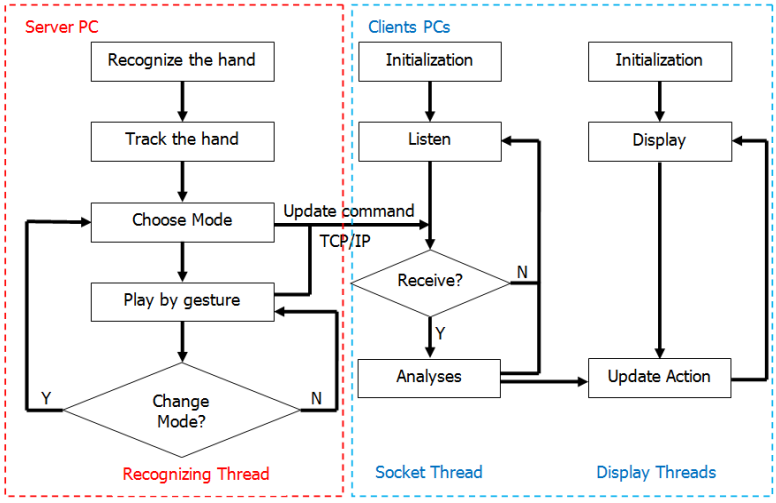

Figure 20. Flow chart of the interactive module

Once the necessary initialization is completed, the server PC awaits the appearance of the manipulator's hand. After recognizing successfully, the server PC enters the Hand-Tracking state and awaits the selection of the operation modes which we define as Scale (SC), Rotate Horizontally (RH), Rotate vertically (RV), Solve Horizontally (SH), Solve vertically (SV).

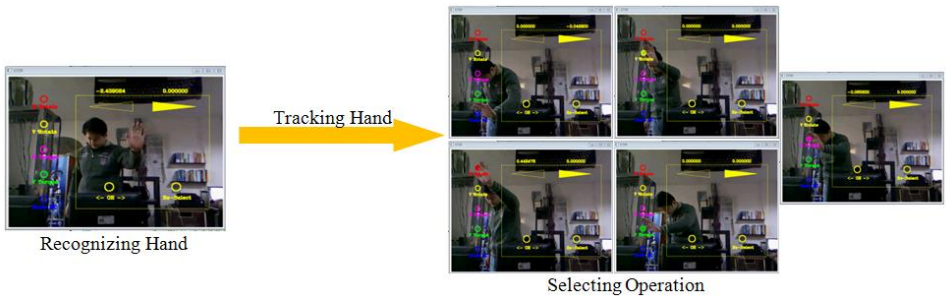

Figure 21. Selecting mode in the interactive module

After the mode selection state in Figure 21, the server PC will then be tracking the movement of the hand and send commands to all the client PCs to update displayed light-field images in real-time. The hand gestures we characterize for the server PC and the corresponding reactions in the client PCs is shown in Table 2.

Table 2 Definitions of the interactive operations in our system

	SC	RH	RV	SH	SV
Move left	NULL	Rotate left as whole	NULL	Rotate left partly	NULL
Move right	NULL	Rotate right as whole	NULL	Rotate right partly	NULL
Move upwards	NULL	NULL	Rotate upwards as whole	NULL	Rotate upwards partly
Move downwards	NULL	NULL	Rotate downwards as whole	NULL	Rotate downwards partly
Move towards along its Z-axis	Zoom in	NULL	NULL	NULL	NULL
Move away along its Z-axis	Zoom out	NULL	NULL	NULL	NULL

In Figure 22a, 22b, we present two typical gestures and their responses in the screen. A yellow rectangle is set for the manipulator to perform the selected operation and several transition states are added to smooth the whole interactive operation. For example, the hand-moving to switch modes should not trigger any interactive reactions in the display. And if one chooses to move right, the draw-back hand-moving from right to left will be counted as invalid to avoid undesired response in the display. Additionally, one can change the move direction or re-select mode by touching the two yellow circles marking "<- OR ->" and "Re-Select", as the Figure 22a, 22b show.

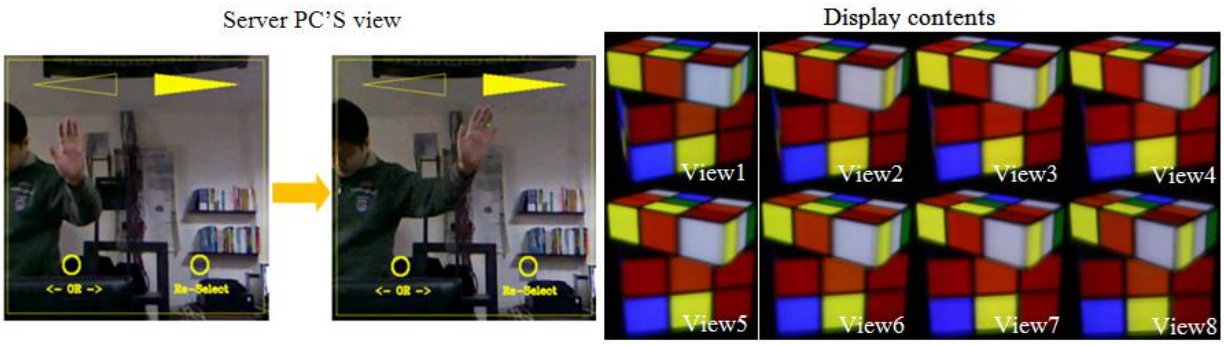

Figure 22a. Move Horizontally

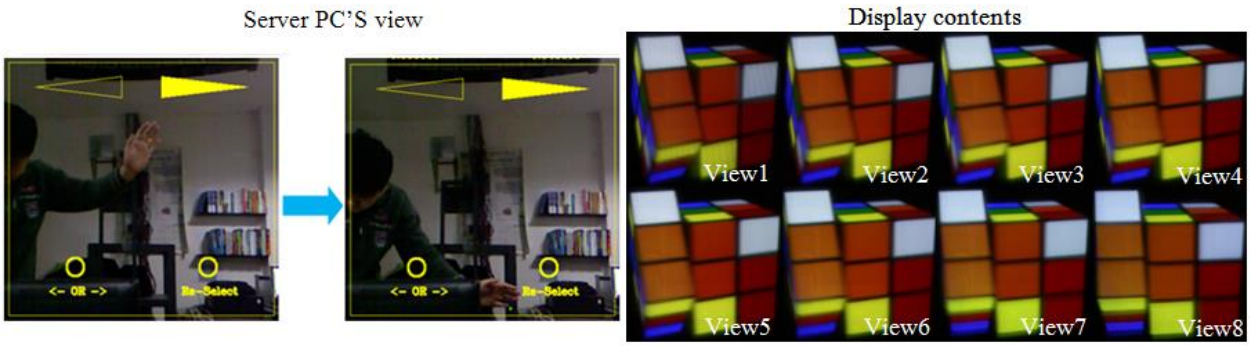

Figure 22b. Move vertically

5. FUTURE WORK

To further improve the quality of the autostereoscopic images to meet commercial requirements, the following steps could be applied to our system.

- **Increase the precision of the lenticular structure**

As we have mentioned in the third paragraph of this paper, the lenticular structure anisotropic screen consists of three components: the lenticular sheet, the glass layer and the diffused substrate. The thickness of the glass layer is determined by the equation (2). The ideal thickness is, however, usually not included in the standardized glass specifications provided by the manufacturers. Therefore, approximation occurs when we choose the glass and the lenticular structure. According to the experiments, the quality of the light-field images is sensitive to the matching of all the variables (r the radius of each lenticular unit, n the material reflection coefficient and e the thickness of the lenticular sheet and glass) in equation (2). And the deviation from the precise value calculated by equation (2) result in weakening of the anisotropic effect, causing blurs in the images. As the system matures in the future and succeeds in mass production, the customized set of the anisotropic screen with all the parameters precisely matching the equation (2) will solve this problem.

- **Customize the light-field display-array**

Being the terminal of the system, the commercial LED projectors are used in the array form to display the light-field images generated by the display card through VGA cable. The function are always inversely proportional to the size of the projector and unfortunately both of these two aspects are emphasized in our display-array. The only possible solution is to develop the customized display equipment with multiview-orient aspects (the efficient cool-down module, the compact layout, the stability and compatibility with various types of video stream formats) strengthened and redundant aspects simplified or deducted. Even, the replacement of the array-form with single light-field display engine module is possible with elaborate optical design to save the cost and make the system more easily to be reproduced.

REFERENCE LINKING

[1] T. Okoshi,[Three-dimensional imaging techniques], New York: Academic Press, (1976).
[2] Barry Blundell and Adam Schwarz,[Volumetric three dimensional display system], NY: John Wiley & Sons, (2000).
[3] D. Gabor, "Holography 1948-1971," Proc. IEEE 60(6), 655 (1978).
[4] Steven Benton and Michael Bove,[Holographic Imaging], NY: Wiley Interscience, (2008).
[5] E. Lueder,[3D Displays], UK: John Wiley & Sons, (2012).
[6] Rolf Hainich and Oliver Bimber,[Displays: Fundamentals & Applications], Boca Raton, FL: AK Peters/CRC Press, (2011).
[7] M Levoy, P Hanraham , "Light field rendering," in Computer Graphics, SIGGRAPH, (1996).

[8] Wojciech Matusik, Hanspeter Pfister, "3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," ACM Trans. Graph. Papers 23(3), 814-824 (2004).
[9] N. Dodgson, "Autostereoscopic 3D Displays," Computer. Papers 38(8), 31 (2005).
[10] G. Favalora, "Volumetric 3D Displays and Application Infrastructure," IEEE Computer. Papers 38(8), 37 (2005).
[11] Downing, E. Hesselink, R. et al, "A three color, solid-state three dimensional display," Science. Papers 273, 1185 (1996).
[12] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, Paul Debevec, "Rendering for an Interactive 360° Light Field Display," in SIGGRAPH 2007 Papers Proceeding, (2007).
[13] Bahram Javidi and Fumio Okano,[Three dimensional television, video, and display technologies], Berlin: Springer, (2011).
[14] J. Geng, "Volumetric 3D Display for Radiation Therapy Planning," IEEE Journal of Display Technology. Papers 4(4), 437 (2008).
[15] J. Geng, "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photon. Papers 3, 128 (2011).
[16] Y Takaki, N Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Optical Express. Papers 18(9), 8824-8835 (2010).
[17] Siegmund Pastoor, Matthias Wöpking, "3-D displays: A review of current technologies," Displays. Papers 17(2), 100-110 (1997).
[18] Zheng GENG, Hui DONG, Tuotuo LI, Mei ZHANG, Zhaoxing ZHANG, Xuan CAO, Renjing PEI, "Multiview three-dimensional display with single projector: Optical design challenges," Optical Express, submitted, (2012).
[19] A. Sullivan, "3 Deep," IEEE Spectrum. Papers 42(4), p. Cover, (2005).
[20] D. Macfarlane, "A Volumetric Three Dimensional Display," Applied Optics. Papers33(31), 7453-7457 (1994).
[21] JUNG-YOUNG SON, BAHRAM JAVIDI, KAE-DAL KWACK, "Methods for Displaying Three-Dimensional Images," Proc. IEEE 94(3), 502-524 (2006).
[22] Jung-Young Son, Javidi, B. ; Yano, S. ; Kyu-Hwan Choi , "Recent Developments in 3-D Imaging Technologies," IEEE Journal of Display Technology. Proc. IEEE 6(10), 394-403 (2010).
[23] Hakan Urey, Kishore V. Chellappan, Erdem Erden, and Phil Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays," Proc. IEEE 99(4), 540-554 (2011).
[24] S. Nagada, "How to Reinforce Perception of Depth in Single Two-Dimensional," Proceedings of the SID, (1984).
[25] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. I. Jsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek and I. Sexton, "An evolutionary and optimized approach on 3D-TV," Proc. of IBC, (2002).
[26] Edward Adelson and James Bergen,[The Plenoptic Function and the Elements of Early Vision in Computational Models of Visual Processing], Cambridge MA, MIT Press, 3-20 (1991).
[27] B. Phong, "Illumination for computer generated pictures," Communications of the ACM. Papers 18(6), 311-317 (1975).
[28] R. Barry Johnson, Gary A Jacobsen, "Advances in lenticular lens arrays for visual display," Proc. of SPIE San Diego 5874, 5874-06 (August 2005).
[29] Intel® Open Source Computer Vision Library, "GetPerspectiveTransform, WarpPerspective," software documentation, Open CV Reference Manual.