Integration of real-time 3D image acquisition and multiview 3D display

Zhaoxing Zhang\textsuperscript{a}, Zheng Geng\textsuperscript{a}, Tuotuo Li\textsuperscript{a}, Wei Li\textsuperscript{b}, Jingyi Wang\textsuperscript{c}, Yongchun Liu\textsuperscript{b}

\textsuperscript{a}State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China; \textsuperscript{b}College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China; \textsuperscript{c}College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT

Seamless integration of 3D acquisition and 3D display systems offers enhanced experience in 3D visualization of the real world objects or scenes. The vivid representation of captured 3D objects displayed on a glasses-free 3D display screen could bring the realistic viewing experience to viewers as if they are viewing real-world scene. Although the technologies in 3D acquisition and 3D display have advanced rapidly in recent years, effort is lacking in studying the seamless integration of these two different aspects of 3D technologies. In this paper, we describe our recent progress on integrating a light-field 3D acquisition system and an autostereoscopic multiview 3D display for real-time light field capture and display. This paper focuses on both the architecture design and the implementation of the hardware and the software of this integrated 3D system. A prototype of the integrated 3D system is built to demonstrate the real-time 3D acquisition and 3D display capability of our proposed system.

Keywords: 3D acquisition, 3D display, integration of 3D acquisition and display, real-time, 3D imaging

1. INTRODUCTION

1.1 The Basic Principle of True 3D Display

True 3D display technology helps viewers gain comprehensive understanding of the high dimensional data or objects[1-9]. As a promising tool to visualize a more realistic world, the true 3D display technology has been a focus in 3D research for over a hundred years[1-18]. A "perfect 3D" display should render virtual objects or real-world scenes in a way that the viewers see them as if through a transparent window[1][17]. 3D images are presented to viewers with correct perspective and depth cues without the need for special glasses or other eyes-tracking wares. Figure 1 illustrates this concept using some images generated in our experiments that show two 3D cartoon characters viewed from different viewpoints (Note: The sizes of objects and screen and locations of viewers are not in proportion to their real-world sizes).

![Figure 1. The "perfect 3D" display](image-url)
1.2 3D Cues provided by display devices

According to physiological optics, 2D images of the world formed on the retina are converted to bio-electrical signals that travel through the nerve to the brain. Each image perceived by our brain is two-dimensional with no three-dimensional information. Nevertheless, we can still accurately perceive through these 2D images the 3D world thanks to the mechanism the brain works with the depth cues presented in these images[6].

3D display devices deceive our brain by providing some of the depth cues, trick us into believing that what we see is 3D. On one hand, various advanced computer graphic technologies have been widely applied in the 2D display fields to provide, out of the monocular image, the depth cues like linear perspective, occlusion, shading and texture as shown in figure 2[17], feeding our brains with the proper images. We can thus perceive 3D on the flat screen when playing 3D video game or using 3D modeling software.

![Figure 2. Illustration of some psychological depth cues from 2D monocular images](image)

On the other hand, binocular images percept at different viewpoints also reveal depth information through convergence of the eyes[17]. As the data-processing capability of the computers advances, more true 3D display devices that make us see two different images simultaneously are now available. So far as our true 3D prototype is concerned, two major binocular depth cues are elaborated in the following part.

1) Binocular disparity (stereo)

Under the assumption of lambertian surface reflection, Stereo image pair acquired with the left and right eyes show resemblance to each other, and relate to each other through binocular disparity map. Considering the convergence ability our brain possess that restrains the ghost-image effect due to the binocular disparity, we therefore gain the final stereo-dominated 3D-perception of these objects or real-world scenes[6].

2) Motion parallax

Motion parallax occurs as the viewing position changes, which gives rise to the motion-parallax-dominated 3D-perception. The further the contents in the real-world scenes, the faster they change positions as the viewer moves and the stronger the view feel this 3D-perception[6].

It is not easy for a display device to generate all the cues mentioned above and generate a "perfect 3D" image for the viewers. The current commercial LCD monitors are good at rendering 2D images which preserver monocular depth cues but fail to deliver the binocular depth cues. Volumetric 3D display featuring strong binocular depth cues usually cannot realize the shading or texture. Some other similar 3D display technologies, such as the stereoscopic display, might incur conflicting
2. 3D DISPLAY AND ACQUISITION TECHNIQUES

2.1 The 3D Display

The 3D display techniques can be generally classified as the geometry-3D-model-based volumetric 3D display and light-field-based multi-view 3D display[6].

1) Volumetric 3D Display

The volumetric 3D display takes advantage of the voxel (the 3D counterpart of the pixel in the 2D display techniques) to reconstruct the virtual objects or real-world scenes. The three-dimensional screens revolved are usually formed by the rotating two-dimensional screens[2-4][13][14], and can present the measurable physical size of the rendered contents. Figure 3 shows the Perspecta volumetric 3D display device produced by the Actuality Systems[14].

![Perspecta volumetric 3D display device produced by the Actuality Systems](image)

As above mentioned, it is difficult to add shading or texture through the current volumetric 3D display techniques. In addition, the huge amount of model data needed to be transferred and processed for displaying large-scale color images still remain challenging for current generation hardware[6].

2) Multi-view 3D Display

The multi-view 3D display techniques basically present images towards different directions for viewers at different viewpoints. Considering the physiological mechanism of the human eyes, the multi-view 3D display techniques simulate the similar physiological process to make the observers believe they "see" the real three-dimensional content. There are several categories of the multi-view 3D display techniques worth discussing:

➤ Stereoscopic 3D display

This technique has been successfully commercialized in the cinema and smart TV due to its excellent 3D experience. Only the lights of corresponding images can pass through the glass lens and reach the eye. The first generation stereoscopic glasses uses red-green filters to selective filter out undesired lights. Current generation stereoscopic glasses use the polarized filters instated of color filter for full color 3D display.(Figure 4)[6].
Anisotropy-film-based autostereoscopic 3D display

The anisotropy-film let through strips of the projected image so the viewers can only see things along the axis connecting the viewing center and the projecting center. The Institute for Creative Technologies (ICT) of the University of Southern California (USC) has invented 3D display devices with a spinning anisotropy-film (Figure 5 left) [2]. The spinning anisotropy screen reflects correct strip of the rendered content to its corresponding direction to display the horizontal 360-degree light-field scenes. The eye-tracking gear can provide the additional vertical motion parallax by tracking the vertical motion of viewers’ eye and correct the perspective accordingly. They update the screen with a flat screen and use projector array in their latest version [4] and the Holografika's series display products HOLOVIZIO™ presents the similar setups (Figure 5 right) [16][19].

Lenticular and micro-lens array autostereoscopic 3D display

Like the anisotropy-film-based 3D display, static or dynamic barriers are applied to give viewers only the needed parts of the whole image. The micro-lens array or lenticular-structure are later put into practice to give a higher resolution [6]. With proper parameters such as the lens per inch (LPI) of the lenticular-structure and the thickness of the diffused substrate [11][17], screens of this type are able to display the large-scale color 3D objects and real-world scenes. Figure 6 shows two typical setups using lenticular and micro-lens array 3D display [17].
2.2 The 3D Acquisition

In our system, light-field imaging techniques are used for 3D acquisition. The light-field image acquisition emphasizes on capturing the light field itself rather than performing geometric 3D reconstruction[6]. Various light-field image acquisition devices are now available including some commercialized products like Lytro[20] and Pelican[21].

![Image](image_url)

Figure 7. Several typical 3D light-field data acquisition array

1) Micro-lens Array Acquisition

The micro-lens array acquisition, also known as the integral imaging, record the light-field data at the two-dimensional image sensor. The acquired light-field data can be used to perform refocusing or viewpoint shift. Figure 7a[6] shows one type of this light-field camera produced by the Lytro[20].

2) Optical Lens Array Acquisition

Similar to the principle of the micro-lens array acquisition techniques, the optical lens array is usually placed in front of the normal lens to acquire the light-field data. Apart from the difference in the light-ray-arrangement order, the optical lens array techniques function almost the same as the micro-lens array techniques. Figure 7b[6][22] and 7d[6][21] demonstrate two typical realizations of the optical lens array techniques.

3) Camera Array Acquisition

As figure 7c shows[6][10][23], large number of cameras in compact forms are arranged to perform the multiview light-field data acquisition. The acquired data can be used to compute the ultra-high resolution image, refocus and shading elimination[6][10][23][24][25].

This article emphasizes on the projection-based multiview display with camera array acquisition and our main contributions that will be elaborated later are that we present an easily-commercialized autostereoscopic multiview 3D display and acquisition system offering large-scale color light-field 3D objects and real-world scenes with improved algorithm and screen[17].
3. SYSTEM OVERVIEW

3.1 System Components

The main components of our system are:

- An anisotropic diffuser screen that primarily reflects vertically;
- PC that controls the display and transmission;
- Display array;
- Acquisition array;

1) Screen

The anisotropic diffuser screen reflects the light primarily in the vertical plane and scarcely in the horizontal plane due to the vertical-diffuse-only (VDO) film we choose. Generally, we put this VDO film towards a transparent plastic board and the projected light images thin highlighted lines after the screen.

Based on what we have mentioned, the imaging process of the screen is analyzed by presenting the ideal light-field of the image surface in one projector[6]:

\[
\ell(x, y, \phi, \varphi) = \begin{cases} 
I(x, y), (\phi(x, y, O_p)) = \theta(x, y, O_p)) \\
0, (\phi(x, y, O_p)) \neq \theta(x, y, O_p)) 
\end{cases} 
\]  
(1)

as \( O_p \) the center of the projector lens, \( \theta(x, y, O_p) \) the direction connecting the pixel positioned at \((x, y)\) with \( O_p \). The expression (1) is reasonable that one cannot see the projected image when facing the projector lens without a diffuse screen.

And now the description of the projected image with the diffuse screen is given[6]:

\[
\ell'(x, y, \phi, \varphi) = \ell(x, y, \phi, \varphi) \otimes \Gamma(x, y, \phi, \varphi) 
\]  
(2)

as \( \Gamma(x, y, \phi, \varphi) \) the diffuse function of the screen and under the ideal condition[6]:

\[
\Gamma(x, y, \phi, \varphi) = \begin{cases} 
1, (x = 0, y = 0, \phi \in (-\frac{\pi}{2}, \frac{\pi}{2}), \varphi \in (-\frac{\pi}{2}, \frac{\pi}{2})) \\
0, |x| + |y| \neq 0 
\end{cases} 
\]  
(3)

The expression (4) can be gained by putting expression (3) into (2) for the whole image[6]:
\[
\sum_{(x,y)} \ell^\prime(x, y, \phi, \varphi) = \sum_{(x,y)} \ell(x, y, \phi, \varphi) \otimes \Gamma(x, y, \phi, \varphi)
\]
\[
= \sum_{\frac{-\pi}{2} \leq \mu, \nu \leq \frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \ell(x, y, \mu, \nu) \cdot \Gamma(x, y, (\phi - \mu), (\varphi - \nu)) d\mu d\nu
\]
\[
= \sum_{\frac{-\pi}{2} \leq \mu, \nu \leq \frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \ell(x, y, \mu, \nu) d\mu d\nu
\]
\[
= \sum_{(x,y)} f(x, y)
\]

This proves the same reflection or transmission light intensity along any direction so the normal image in the diffuse screen can be seen.

As for the VDO film in our screen, the expression (3) can be deduced to:

\[
\Gamma(x, y, \phi, \varphi) = \begin{cases} 
\Gamma_{\sigma_\phi, \sigma_\varphi}(\phi, \varphi), (x = 0, y = 0, \phi \in (-\frac{\pi}{2}, \frac{\pi}{2}), \varphi \in (-\frac{\pi}{2}, \frac{\pi}{2})) \\
0, (|x| + |y| \neq 0)
\end{cases}
\]

with \(\sigma_\phi, \sigma_\varphi\) two independent probability distribution function and \(\Gamma_{\sigma_\phi, \sigma_\varphi}(\phi, \varphi)\) their joint probability distribution function. The VDO film gives a maximum 60-degree vertical diffuse and less than 1-degree horizontal diffuse[2][6] so we let:

\[
\begin{align*}
\sigma_\phi &= 1 \text{deg} \\
\sigma_\varphi &= 60 \text{deg}
\end{align*}
\]

and the experimental intensity distribution of the VDO film can be shown in figure 8[6]:

![Image](image.png)

Figure 8. Horizontal and vertical transmission light intensity distribution of the VDO film (middle: micro structure of the VDO film; right: VDO light intensity distribution result)

So the conclusion can be made that the VDO film is horizontal-sensitive and the film can be treated as the normal diffuser within the 1-degree angular interval[3-6].

Then the expression (4) can be rewritten using (5) and (6):
\[ \sum_{(x,y)} \ell'(x, y, \phi, \varphi) = \sum_{(x,y)} \ell(x,y,\phi,\varphi) \otimes \Gamma(x,y,\phi,\varphi) \]
\[ = \sum_{(x,y)} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \ell(x,y,\theta(x,y,O_p) \mid \phi, \varphi) \cdot \Gamma_{\text{deg,60deg}}(x,y,(\phi - \theta(x,y,O_p) \mid \phi),(\varphi - \theta(x,y,O_p) \mid \varphi)) \, d\mu d\nu \]
\[ = \sum_{(x,y)} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} I(x,y) \cdot \Gamma_{\text{deg,60deg}}((\phi - \theta(x,y,O_p) \mid \phi),(\varphi - \theta(x,y,O_p) \mid \varphi)) \, d\mu d\nu \]
\[ = \sum_{(x,y)} I(x,y) \cdot \Gamma_{\text{deg,60deg}}((\phi - \theta(x,y,O_p) \mid \phi),(\varphi - \theta(x,y,O_p) \mid \varphi)) \]

This shows the VDO film modulation imposed upon the image that multiple projectors will generate at the observing point \(O_x\). And the superposition of all the modulated projected images can be generalized as:

\[ I_{\text{out}} = \sum_{i_{\text{projector}}} \sum_{(x,y)} I(x,y) \cdot \Gamma_{\text{deg,60deg}}((\phi - \theta(x,y,O_p) \mid \phi),(\varphi - \theta(x,y,O_p) \mid \varphi)) \]
\[ = \sum_{i_{\text{projector}}} \sum_{(x,y)} I(x,y) \cdot \Gamma_{\text{deg,60deg}}((\phi - \theta(x,y,O_p) \mid \phi),(\varphi - \theta(x,y,O_p) \mid \varphi)) \]

Therefore, we can select the right perspective of the rendered objects or real-world scenes for the viewer by controlling the projected contents in each of these projectors.

2) Display Array

The display array is designed in a compact way so the prototype turns more mobile and exhibition-friendly. We use 54 mini-type Dell® M115 LED projectors in the array and wide angle lens is attached to each projector to shorten the light path. A dedicated connecting socket is designed using Solidworks to attach the projector with the wide angle lens, as figure 9 shows.

![Figure 9. Projector with wide angle lens (left: Solidworks model)](image)

3) Rendering hardware

As shown in figure 10, our rendering hardware consists of a workstation with three multi-head display cards and 18 video splitters. Each of these graphic cards on the workstation can generate up to six independent video outputs, the resolution of each output is 1280*2400, which can be further split to three 1280*800 video outputs, thus each graphic card can be used to drive up to 18 projectors.
4) Acquisition array

The acquisition hardware architecture is designed as shown in Figure 11 and 12. We can either transmit from each display unit the captured data to the workstation for further processing (the Camera-RasPI-PC mode) or to the corresponding display unit through the video interface the acquisition unit provides (the Camera-RasPI-Projector mode).

We use the low-priced Raspberry Pi® (RasPI) board as the acquisition unit. The equipped camera supports up to 1080p HD video recording and maximum 2592*1944 resolution.

![Figure 10. Hardware system of the display array sub-system (not all the identical splitter and projector setups are shown)](image)

![Figure 11. Hardware system of the acquisition array in Camera-RasPI-PC mode (not all the identical RasPI and camera setups are shown)](image)

![Figure 12. Hardware system of the acquisition array in Camera-RasPI-Projector mode (not all the identical RasPI, camera and switch setups are shown)](image)

The acquisition unit also features a compact and expandable form so we can arrange more than one acquisition arrays horizontally or vertically if needed. Figure 13 shows the design of our acquisition array:
5) Software architecture

Our system firstly calibrates all the display units by means of looking for the homography matrix (HM) between the standard chessboard calibration picture (CCP) and the projected CCP and the HM between the standard CCP and the screen CCP. These two HMs help realize the geometric calibration in our system. In the rendering mode, our system could use the dll-injection[26] to acquire the light-field data of the rendered object[24] and the high level shader language (HLSL) scripts are used for vertex and texture matrix transforms so the correct 3D light-field image can be shown in the screen[6]. The capture module takes the captured data from the acquisition array and display the 3D light-field scene with rearrangement algorithm. Figure 14 gives an overview of the primary procedures:

We use the Microsoft® Direct3D (D3D) to realize the multi-monitor rendering process. And the Raspbian OS are used (also officially recommended by the manufacturer of the Raspberry PI) in the acquisition array to complete the light-field capture work.
3.2 System functions

Almost all the steps leading to the final rendered 3D light-field image have been elaborated and we will present a detailing discussion about those that we contribute.

1) Calibration

Projected images from multi-projectors positioned differently will give rise to perspective-warped quadrilaterals instead of uniform rectangles on the display plane[17]. The rotation and translation of each projector lead to the warped projected image. Therefore the HM between the standard CCP and the warped CCP recording the rotation and translation are needed to perform the reverse-rotation and reverse-translation operation[6]. The expression (9) shows how we get the HM:

\[ HM(x, y) = HM_1(x, y, u, v) \cdot HM_2(a, b, u, v) \cdot HM_3(x, y, u, v)^{-1} \]  

Let x, y be the coordinates of the CCP to be projected, u, v of the captured CCP and a, b of the CCP suited to the screen and the HM(x, y) is gained for each projector. In the Direct3D, the post-multiplication in expression (9) is applied due to its way to store the elements in the matrix.

The OpenCV library includes functions like WarpPerspective and GetPerspectiveTransform[27] to perform the calibration[17]. The finding-HM method used in this paper proves a faster and more efficient way to calibrate all the projectors in our system. Figure 15 shows the calibration procedures:

![Calibration procedures](image)

Figure 15. Calibration procedures in our system

The key step among these steps is to unify all the coordinate systems to calculate the correct HMs. Figure 16 shows the change of the coordinates system and figure 17 shows the warped and de-warped CCP displayed in the screen.

![Finding HMs](image)

Figure 16. Finding the HMs in the geometric calibration
Figure 17. Warped (left four small images before calibration) and de-warped (right image after calibration) CCP displayed in the screen

2) Rendering

Unlike the old method rendering all the perspective images viewed from their corresponding viewpoints[17], direct manipulation of the rendering parameters of the virtual object can be made in the pipeline through the HLSL scripts. Based upon the expression (8) and previous discussions, the vertical strip within the 1-degree angular interval of the VDO film can be conditioned as:

$$\phi \mid_{\Phi_{\phi}} - \theta(x, y, O_{\phi}) \mid_{\phi} \leq 1 \text{deg}$$

(10)

the viewer at $O_{c}$ will see all the projected pixels in the same direction. Considering only the yoz plane in the rendering space, as the left part of the figure 18 shows:

![Diagram](image)

Figure 18. The yoz (left) and xoz space we render object

Within the plane defined by expression (10), two points $P_{1}, P_{4}$ in the virtual objects are traced from where the viewer at $O_{c}$ and the cross-points $P_{2}, P_{3}$ with the screen are the rendered points we should draw to show $P_{1}, P_{4}$ respectively. And expression (11) can be deduced:

$$\tan \angle P_{1}O_{c}P_{3} = \frac{y_{N} - y_{0}}{z_{N} - z_{0}} = \frac{y_{N} - y_{0}}{z_{N} - z_{0}} = \overline{y}_{P_{1}}$$

$$\tan \angle P_{2}O_{c}P_{5} = \frac{y_{N} - y_{0}}{z_{N} - z_{0}} = \frac{y_{N} - y_{0}}{z_{N} - z_{0}} = \overline{y}_{P_{4}}$$

(11)

The rendered image projected will hold a different expression:
\[
\tan \angle P_{p_{e}}P_{e} = \frac{y_{R} - y_{O_{e}}}{z_{R} - z_{O_{e}}} = \frac{y_{R} - y_{O_{e}} - (y_{O_{e}} - y_{O_{e}})}{z_{R} - z_{O_{e}}} = \frac{y_{R}}{z_{R} - z_{O_{e}}} \quad (12)
\]

\[
\tan \angle P_{O_{e}}P_{e} = \frac{y_{R} - y_{O_{e}}}{z_{R} - z_{O_{e}}} = \frac{y_{R} - y_{O_{e}} - (y_{O_{e}} - y_{O_{e}})}{z_{R} - z_{O_{e}}} = \frac{y_{R}}{z_{R} - z_{O_{e}}} \quad (12)
\]

The discrepancy between the projecting side and the viewing side in terms of the same point's view angle can be removed by:

\[
y_{O_{e}} = y_{O_{e}} \quad (13)
\]

Such design can safely put expression (12) into (13) and relate the projecting side with the viewing side in the plane defined by expression (10) as[2][6]:

\[
y_{\text{observed}} = \left( \frac{z_{p} - z_{O_{e}}}{z_{p} - z_{O_{e}}} \right) \cdot y_{\text{projected}} = \kappa \cdot y_{\text{projected}} \quad (14)
\]

Seen from top of all the system, the x coordinates of these points within the 1-degree angular interval is gained. Free from the y coordinates discrepancy we have discussed, the expression (15) can be deduced according to the right part of figure 18:

\[
x_{\text{observed}} = x_{\text{projected}} = \frac{x_{p} - x_{O_{e}}}{z_{p} - z_{O_{e}}} \quad (15)
\]

With the help of expression (14) and (15), we mainly manipulate the x-y-z coordinate in the HLSL scripts to achieve the rendering. As for the captured images, the rearrangement algorithm is applied in the same way. The linear interpolation might be used if the captured light-field is not dense enough. Apart from what we have processed, we also need to consider the face culling, depth test and set these options correctly in the program to get the final 3D image.

3) Acquisition

As mentioned before, the low-priced RasPi is selected as the acquisition unit in our acquisition array to get the light-field data and feed these data to the display end. The camera features an ov5647 sensor connected to the camera serial interface in the RasPi board through a 15 pin ribbon cable[28]. The 720p video frame rate is up to 30 frame per second (fps) using the YUV 4:2:0 video format.

In the Camera-RasPi-Projector mode described in figure 12, we need to port the calibration process from PC to the RasPi board and it takes a lot more time to finish due to the limited computing capability of the RasPi (single CPU with 700MHz, 512MB memory). As for the Camera-RasPi-PC mode, the workstation can handle almost all the computations. However, it still takes time to finish all the viewpoints' images. The overhead increases as we expand our system by adding more arrays. Nevertheless, we try to improve the rectification process by using the look-up-table LUT method[17] and adapt it to match the YUV video format to accelerate the rectification process in the RasPi. The expression (16) presents the adapted algorithm.

\[
Y_{\text{calib}} = \begin{cases} 
    f(M_{\text{calib}}(Y_{r}, Y_{i})) = LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame = 1}) \\
    LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame > 1})
\end{cases} 
\]

\[
U_{\text{calib}} = \begin{cases} 
    f(M_{\text{calib}}(U_{r}, U_{i})) = LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame = 1}) \\
    LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame > 1})
\end{cases} 
\]

\[
V_{\text{calib}} = \begin{cases} 
    f(M_{\text{calib}}(V_{r}, V_{i})) = LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame = 1}) \\
    LUT_{i,j,0}(0 \leq i \leq 1280, 0 \leq j \leq 800, \text{frame > 1})
\end{cases} 
\]


And table 1 summarizes a simple result of the contrast experiment in the rectification part performed by the RasPI:

<table>
<thead>
<tr>
<th>OpenCV(embedded version)</th>
<th>LUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>fps(1280*800 resolution)</td>
<td>&lt;1.0</td>
</tr>
</tbody>
</table>

Considering all the other time-consuming algorithms the whole system needs, the display rate will be dragged down to less than 1 fps so we focus on static images display for the time being.

4) Transmission

In either mode mentioned above, the efficient transmission method is required. Generally, the send buffer and receive buffer are set to handle the image data in a non-block way. In addition, the finite-state-machine (FSM) is applied to lower the data loss ratio as much as possible under the UDP. And the synchronize thread is also designed to ensure harmony between different software modules. Figure 19 shows the flow chart of the system:

Figure 19. Flow chart of the transmission modules in our system

Noticeably, every acquisition unit needs to send part of its light-field data and receive parts of all the other units’ light-field data under the Camera-RasPI-Projector mode. So several gigabit switches are applied to connect all the acquisition units together to the workstation.

4. EXPERIMENT RESULTS OF THE SYSTEM

4.1 Rendered Models

We first give two groups of models rendered at three different angular intervals (the 0.3-degree-β section, 0.6-degree-β section and 0.9-degree-β section). Limited by the space, we evenly select 10 images out of the total 54 ones observed at each angular-interval section.
The larger angular $\beta$ can give rise to stronger horizontal parallax only (HPO) effect as we can sense more obvious perspective difference from the 0.9-degree-$\beta$ section in figure 20 and 21. The enlargements of the same part in each model are also compared to show that the water plant micro structure in figure 20 becomes blurred as the angular interval $\beta$ increases while the tentacles of the dragons in figure 21 seems not affected by this.

4.2 Captured Scenes

We then present two groups of real-world scenes captured by our acquisition array and transmitted to the display end with the same angular interval setup. Still, only images evenly taken at 10 viewpoint positions out of the total 54 ones in each angular-interval section are shown.
It can also be concluded that larger angle interval generates more evident HPO effect as the first and the tenth images of the mannequin head in figure 23 differ more in the 0.9-degree-\(\beta\) section than those in the 0.3-degree-\(\beta\). The smaller of the angular interval will, however, restrain the ghosting effect, as the enlarged front leaf tip in figure 24 gets more blurred in 0.9-degree-\(\beta\) section than that in 0.3-degree-\(\beta\) section.

### 4.3 Summary

Based on the elaboration revolved around the angular interval above, it is known that stronger HPO can be achieved for some types of display contents by applying larger angular interval and we try to categorize the contents displayed by our system in table 2:

<p>| Table 2. Comparison of subjective 3D image quality at different angular intervals for different contents | 3D display of one potted plant captured at three different angular interval (\beta) | 3D display of one mannequin head captured at three different angular interval (\beta) |</p>
<table>
<thead>
<tr>
<th>Contents</th>
<th>$\beta =0.3$ degree</th>
<th>$\beta =0.6$ degree</th>
<th>$\beta =0.9$ degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light reflection boundary on the tips of the leaves</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Mannequin head lips</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tentacles of the dragon without texture</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Water plants with fine micro-structure</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

It's difficult to give a perfect subjective evaluation of the image[29] so we simplify this by setting the subjective quality of the images at the 0.3-degree-ß section as the unit one. And we add one in table 2 if the image becomes blurry. Table 2 tells that different types of contents displayed require different angular intervals for suitable 3D perception. Those rich in details such as the water plant micro structure call for smaller angular interval while sparsely-textured contents like the dragon tentacles seems to be free from this limit and can be applied with larger angular interval to gain stronger HPO-perception. Conclusively, we need densely-angular-interval-sampled light-field data for the complex objects or real-world scenes that might be full of details we are interested in and prepare light-field data angular-spaced more largely for the simple less-textured objects or real-world scenes to strengthen their HPO effect at less cost of lowering their 3D image quality.

5. FUTURE WORK

To improve our system, the following steps could be considered in the future:

- **Customized hardware for the processing of the captured images**
  
  According to figure 18, light-field data can be exchanged between different acquisition units. It takes $C_N^2$ of times for the complete data exchanging process in one frame in an N-node acquisition array. The display rate might thus be dragged down to less than 1 fps at the 720p resolution. Customized hardware with enough in-board memory allocated for each unit's light-field data could be designed to perform the exchanging process and other time-consuming algorithms in a more parallel way. And the ideal number of data exchanging could be lowered from $C_N^2$ to N as figure 24 shows. In addition, the customized hardware platform could possibly edge out the low-frequency and not GPU-friendly RasPI board in terms of the computation capability.

![Figure 24. Current and planned-improved (right) acquisition topology structure](image)

- **Content-oriented 3D display**
  
  As the conclusion in the experiment results paragraph, there are different angular interval standards for different types of contents displayed. Smaller angular interval means finer details with weaker HPO-perception and vice versa. Therefore, we could render and transmit light-field data with larger angular interval in the region-of-less-detail-dominated (ROLD-dominated) frames to gain stronger HPO-perception. Conversely, we could apply smaller angular
interval for the region-of-more-detail-dominated (ROMD-dominated) frames to give us more details. The ROLD-dominated or ROMD-dominated property could be set for each frame automatically or manually.

6. ACKNOWLEDGMENT

This work has been supported by the National High-tech R&D Program (863 Program) of Institute of Automation, Chinese Academy of Sciences (CASIA), grant 2012AA011903.

REFERENCES


