A penetrable interactive 3D display based on motion recognition

(Invited Paper)

Chen Su (苏 忱)[1], Xinxing Xia (夏新星)[1], Haifeng Li (李海峰)[1*], Xu Liu (刘 旭)[1], Cuifang Kuang (匡翠方)[1], Jun Xia (夏 军)[2], and Baoping Wang (王保平)[2]

[1]*State Key Laboratory of Modern Optical Instrumentation, Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China*

[2]*Southeast University, Nanjing, Jiangsu 210096, China*

[*]*Corresponding author: lihaifeng@zju.edu.cn*

Received March 31, 2014; accepted April 23, 2014; posted online May 30, 2014

Based on light field reconstruction and motion recognition technique, a penetrable interactive floating 3D display system is proposed. The system consists of a high-frame-rate projector, a flat directional diffusing screen, a high-speed data transmission module, and a Kinect somatosensory device. The floating occlusion-correct 3D image could rotate around some axis at different speeds according to user's hand motion. Eight motion directions and speed are detected accurately, and the prototype system operates efficiently with a recognition accuracy of 90% on average.

OCIS codes: 120.2040, 100.6890.
doi: 10.3788/COL201412.060007.

With the progress of computer and optoelectronics techniques, three-dimensional (3D) display has experienced an unprecedented development in recent years[1,2]. It has been the dream of human beings to create a 3D image in space just like a real object set before the observer. It can also be described as: the glasses-free display scene can be watched around in space with correct field of view and correct spatial occlusion effect, and observers can interact with the 3D image as if operating a real object. Currently, there are several methods to achieve the glasses-free 3D display. Take the volumetric display as an example, it is a spatial addressing display, which could only reconstruct voxels' location information apart from the voxels' light angular distribution[3−5]. As a result, the volumetric displaying scene has no occlusion effect. The representative method to achieve occlusion-correct 3D display is to reconstruct the light field of the 3D scene. Light field reconstruction method simulates the way in which real 3D scenes emit rays and display corresponding images for different views. So all observers around could observe 3D images at their own positions.

Light field display has attracted lots of attention in the last few years. Based on Perspecta 3D System, Cossairt *et al.*[6] limited the screen's light diffusing angle and realized a 198-view occlusion-capable volumetric 3D display. Yendo *et al.*[7] achieved a 360-degree dynamic color 3D display, who utilized the slowly rotating cylindrical parallax barrier combining with a rapidly counter-rotating linear LED array. Jones *et al.* employed a high-speed projector and a rotating holographic screen to realize an interactive monochrome 360-degree light field display[8].

However, all above-mentioned light field 3D displays are not penetrable, which means that the screen separates the observer from the display region, and it would lead to an unnatural interaction with the 3D scene. Some research works on the floating 3D display have recently been proposed. Takaki and Uchida proposed to use scanning multi-view method to create a floating 3D display, and the utilized screen made the projector image to the viewer's position[9]. Our research group employed a LED-based high-frame-rate projector and a flat light field scanning screen to create the light field of real 3D scene in the air above the screen[10]. Although these methods can display a floating image, users still cannot interact with it.

Based on the previous research work on the 360-degree light field 3D display[11], we add an interactive device to this system and implement a novel motion-based interactive penetrable 3D display system.

The proposed 3D display system chiefly includes a floating 3D display module, a motion recognition module, a high-speed data transmission module, and a computer as shown in Fig. 1. According to the hand motion information captured and analyzed by the motion recognition module, the corresponding image package would be transmitted to the projector's buffer through the transmission module, and the corresponding 3D image could be displayed as if floating upon the screen. As a result, the interaction with the displayed 3D scene is realized. It will be more immersive and interesting when observers penetrate and interact with the reconstructed floating 3D scenes.

The module is mainly comprised of a high-frame-rate projector, a flat directional diffusing screen, and a revolving mechanism. The flat screen here is a circular reflective directional-diffusing screen. Benefiting from the microstructure on the screen, the screen can deflect the normal incident light to a certain tilted angle, which horizontally (or circularly) reflects light in a very small angular range and diffuses light in a large angle vertically.

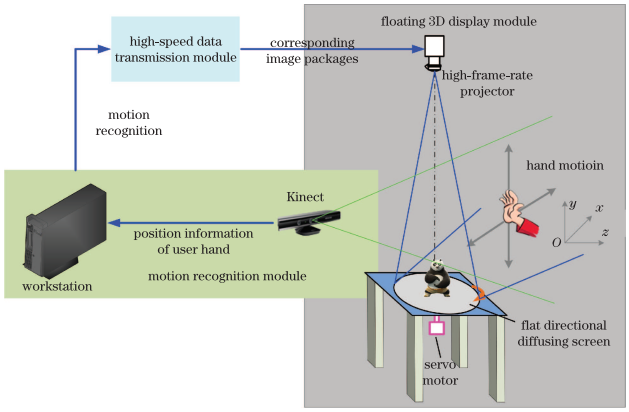


Fig. 1. Proposed interactive 3D display system.

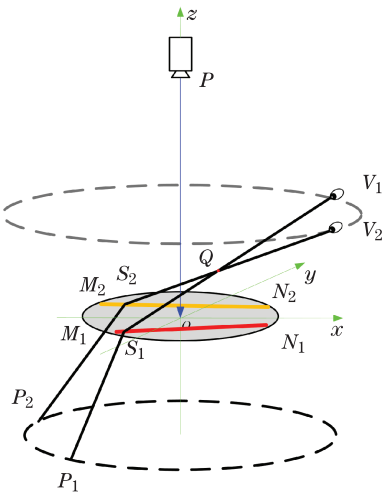


Fig. 2. Mapping relationship.

When the revolving screen is irradiated by white light, a narrow strip can be watched to sweep across the screen. This process is defined as "partly sector scanning mode", and the scanning center can be calculated according to the screen's particular reflection characteristics. Based on the scanning mode, the correct horizontal light field could be reconstructed by projecting synchronizing matching images onto the high-speed revolving flat directional diffusing screen to achieve 3D display floating in the air. It needs pointing out that the images are synthetic for different views.

From the perspective of light field, the basic display principle is illustrated in Fig. 2. When the projector projects images on the rotating screen, it is equivalent to a huge number of mirror projectors P_n spinning with a circle under the screen. When the eye located at the position of V_1 observes spatial point Q, only the ray P_1S_1 projected by the projector P_1 could reconstruct the ray QV_1. In other words, the intensity information of Q at this view should be recorded in the corresponding pixel at S_1. Similarly, for an arbitrary viewpoint, there is a corresponding projector projecting images to reconstruct point Q. When the number of mirror projectors is large enough, it can be assumed that the 360-degree light field of Q is reconstructed. So the mapping relationship between the original 3D model and the projection images can be established, which can be defined as

$$\sum_{\text{img}} = f(\boldsymbol{\Theta}), \tag{1}$$

where $\boldsymbol{\Theta}$ is the spatial matrix of the original 3D model, $\sum_{\text{img}}$ is the group of projection images by which a static 3D image can be displayed, and $f(\ )$ indicates the mapping relationship between the projection images and spatial points, which has been discussed in detail in Ref. [11].

The Kinect (Microsoft Corporation, USA) is utilized as the motion capture device. Benefiting from the infrared ray sensors and cameras, Kinect can track the user's key skeleton points and return their 3D position information (x, y, and z).

In the motion recognition program, the 4-dimension array hand $[x][y][z][t]$ is defined to store the coordinate of the hand, which indicates the hand's spatial position at some time point. Obviously, t is related with sampling time and linealy increasing strictly. As shown in Fig. 3, when the user waves his hand from A to B, there are many sampling points t_i in a selected period of time from t_{start} to t_{end}. Now considering two neighboring sampling nodes (P_i and P_{i+1}), the vector can be decomposed to δx_i, δy_i, and δz_i (δz_i can be dismissed for simplification) orthogonally. So the inclination angle θ_i ($0 \leqslant \theta_i < 2\pi$) could be expressed as a trigonometric function. Obviously, θ_i is not stable, but it can be classified to several base directions which have been set in advance. For all the sampling points from t_{start} to t_{end}, if the following equation is satisfied constantly, the motion meaning would be regressed as the corresponding base direction:

$$|\theta_i - \alpha| < \varepsilon, \tag{2}$$

where α is a base direction angle, and ε indicates a tolerance value which is set by users according to the specific circumstances.We could also speculate the motion speed by analyzing the value of δx_i or δy_i, and the sampling time interval.

As shown in the $x-y$ coordinates system in Fig. 3, if α=0, the motion would be defined as "wave right", and if α=$3\pi/4$, the motion direction would be bottom left.

The relationship between motion and rotating trajectory is illustrated in Fig. 4. The red and blue vectors indicate the motion trajectory and the rotating trail of the 3D image respectively.

For example, the display system will show the 3D image

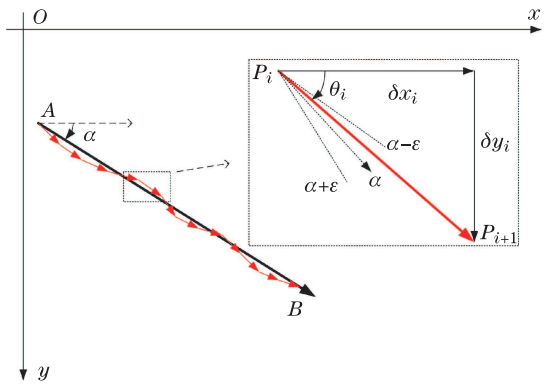


Fig. 3. Motion recognition algorithm.

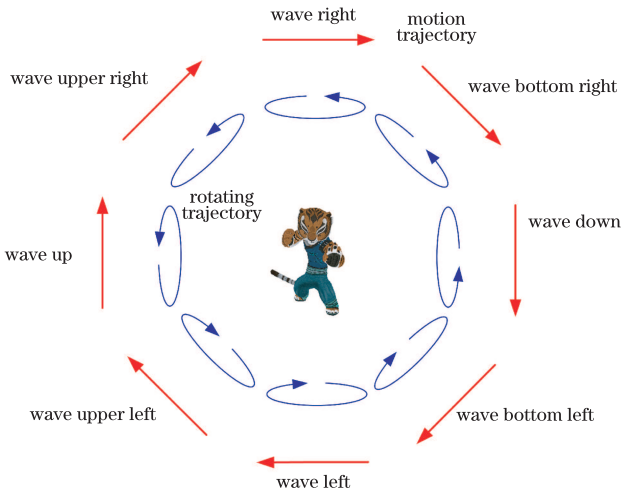


Fig. 4. Relationship between motion and rotating trajectory.

revolving anticlockwise along the z axis after the motion of "wave right". This equals that a sequence of static 3D images with different spin angles are displayed consecutively.

Based on Eq. (1), this process can be defined as

$$\sum_{\text{img}} = f[R^{\theta}_{axis}(\mathbf{\Theta})]. \tag{3}$$

Here, R^{θ}_{axis} rotates the original model's spatial matrix $\mathbf{\Theta}$ in θ degree with axis. It can be defined as follows. Given the spin axis and spin angle, a point $Q(x, y, z)$ in $\mathbf{\Theta}$ will be converted to $Q'(x', y', z')$ in accordance with

$$[x', y', z']^{\mathrm{T}} = \mathrm{E}^{\mathrm{i}\alpha}\mathrm{E}^{\mathrm{j}\beta}\mathrm{E}^{\mathrm{k}\gamma}[x, y, z]^{\mathrm{T}} \tag{4}$$

$\mathrm{E}^{\mathrm{i}\alpha}$, $\mathrm{E}^{\mathrm{j}\beta}$, and $\mathrm{E}^{\mathrm{k}\gamma}$ are the rotation matrix with x, y, and z spin axis respectively, and α, β, and γ indicate the spin angle. They can be described as

$$\mathrm{E}^{\mathrm{i}\alpha} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \mathrm{E}^{\mathrm{i}\beta} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\cdot\mathrm{E}^{\mathrm{k}\gamma} = \begin{bmatrix} \cos\gamma & -sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

The whole model will be rotated when all points are operated by the matrices.

As the example mentioned above, when we set K (K is an integer) static 3D images in one circle, the axis will be set as z axis, so β=0, γ=0, α=$i\times360°/K$ (i=1$\cdots K-1$). K groups of projection images would be packed compressively and stored in a hard disk awaiting the call of motion recognition module.

A high-speed data transmission module is necessary for real-time motion-based interaction. In our system, we utilize the PCI-E Bus to improve the transmission speed significantly. The structure and process of the high speed data transmission module are shown in Fig. 5. The data of three-channel images is transmitted to the corresponding FPGA by coaxial connectors and then distributed to the D4100 control panels which drive DMD chips to display images. When a useful motion happens, the corresponding image packages will be transmitted to the buffer of projector and the rotated 3D scenes are displayed. So the interaction is realized finally.

A penetrable interactive 3D display system based on motion recognition is developed in the experiment. The prototype configuration is illustrated in Fig. 6. Three-chip DMD-based color high-frame-rate projector is implemented with RGB LEDs (PhlatLight PT54 of Luminus Devices) as the light source. The spatial light modulator of high-frame-rate projector used is Discovery 4100 Kit purchased from Texas Instruments, which can display at most 32552 single-bit frames per second with the resolution of 1024×768. In the projector three chips of DMD are utilized to display three-channel images synchronously. The screen, which reflects normal incident ray with 45-degree, is mounted on the revolving mechanism. Two mirrors are utilized to fold the image from the projector to the screen perpendicularly. The Kinect device is placed 0.8 m away from the display system to recognize observer's motion.

In the experiment, the raw spatial position data has been mapped into the plane perpendicular to the optical axis of the Kinect device, and converted to the pixel information in the screen. We defined 8 base motion directions as shown in Fig. 3. Figures 7(a) and (b) indicate the user right hand's spatial coordinate when he

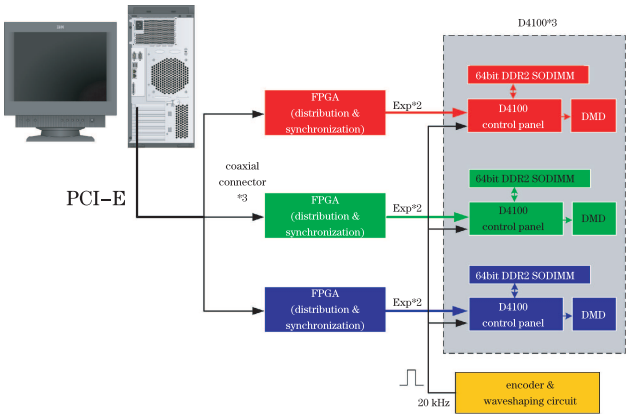


Fig. 5. Process of the high speed data transmission.

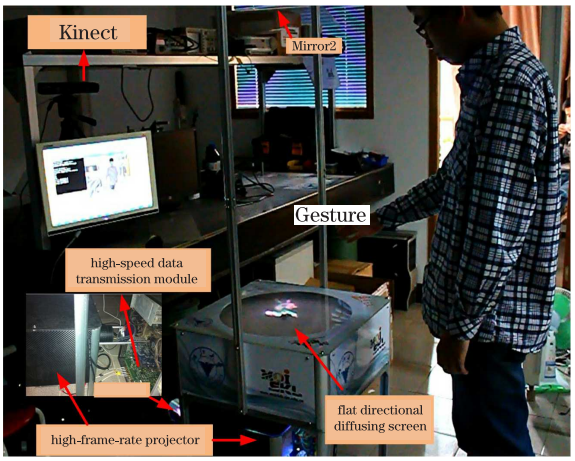


Fig. 6. Prototype of interactive 3D display system.

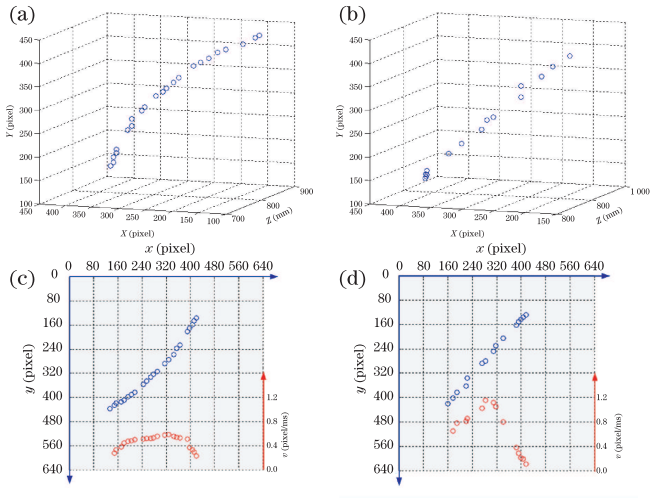
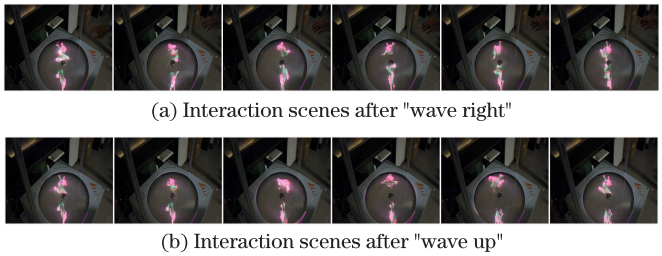


Fig. 7. User hand's spatial coordinate while waving hand from upper right to bottom left (a) more slowly and (b) more rapidly. Hand's $x-y$ coordinates (c) in slower motion speed and (d) in faster motion speed.



(a) Interaction scenes after "wave right"

(b) Interaction scenes after "wave up"

Fig. 8. Images captured from the interaction video we recorded.

waves hand from upper right to bottom left slowly and rapidly respectively. The Kinect's sampling time interval is about 30 ms (not really stable) and the hand-waving distance is about 1.0 m. In Figs. 7(c) and (d), the blue points show the right hand's x-y coordinates in slower (about 1.4 m/s) and faster (about 2.3 m/s) motion speeds respectively. The red points indicate the instantaneous velocity, and the maximum values are 1.15 pixels/ms and 0.58 pixels/ms, respectively. As a result, the average motion speed could be figured out, corresponding to the revolving speed of the 3D scene. The prototype system operates efficiently with a mean recognition accuracy of 90%.

To make the interaction closer to the reality, motion recognition is triggered when one's hand (left or right) penetrates to the 3D display area. According to different motions, different projection image packages have been generated previously and stored compressively in the hard disk awaiting the call of motion recognition module. Instead of generating the projection images dynamically, this way of look-up table increases response speed much further. In Fig. 8, two interaction scenes are shown by the images captured from the interaction video. In the video, after the user waved his hand, the 3D image rotated correspondingly.

For a static 3D image, 600 is chosen as the number of projection images to reduce flicker. The rotating speed of the screen is 1800 rpm, so the refresh frequency of the static 3D scene is 30 Hz, and the display time for each projection image is 55.5 μs. For an interactive dynamic 3D image, the spin angle between two continuous positions is set as 36°, so K=10. In consideration of the starting and ending positions, the number of projection images is 6600×3 (3 indicates three channels), and the compression ratio of the images package is about 10%. The response time between a useful motion and display is about 0.5 second, most of which is spent opening up the transmission module. The frame frequency of the dynamic 3D scene is about 5 fps.

In this letter, we present a penetrable interactive 3D display system based on motion recognition, which is suitable for users to interact with the floating virtual 3D images. The motion recognition module identifies motion based on hand waving and calls the high-speed data transmission module to transmit the image which is package generated and stored in the hard disk previously to the 3D display system. The response time and the frame frequency of dynamic 3D scene are limited by the data transmission rate and image package's size. Hence, it is our future work to improve the interactive performance by optimizing the transmission module from serial to parallel and increasing the compression ratio of the images package. Due to the limitation of Kinect's FOV, the users can just interact within an area. So the 360-degree multi-user interactive 3D display system is also considered.

This work was supported by the National Basic Research Program of China (973 Program) (No. 2013CB328806), the National High Technology Research and Development Program of China (863 Program) (No. 2012AA011902), the National Natural Science Foundation of China (No. 61177015), and the Research Funds for the Central Universities of China (No. 2012XZZX013).

References

1. J. Hong, Y. Kim, H. J. Choi, J. Hahn, J. H. Park. H. Kim, S. W. Min, N. Chen, and B. Lee, Appl. Opt. **50,** H87 (2011).
2. N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, IEEE Trans. **57,** 362 (2011).
3. G. E. Favalora, Computer **38,** 8 (2005).
4. G. E. Favalora, Proc. SPIE **4712,** 300 (2002).
5. X. Xie, X. Liu, and Y. Lin, J. Opt. A: Pure Appl. Opt. **11,** 045707 (2009).
6. O. S. Cossairt, J. Napoli, S. L. Hill, R. K. Dorval, and G. E. Favalora, Appl. Opt. **46,** 1244 (2007).
7. T. Yendo, T. Fujii, M. Tanimoto, and M. Tehrani, J. Vis. Commun. Image. **21,** 586 (2010).
8. A. Jones, I. McDowall, H. Yamada, M. Bolas and P. Debevec, ACM Trans. on Graphics **26,** 40 (2007).
9. Y. Takaki and S. Uchida, Opt. Express **20,** 8848 (2012).
10. X. Xia, C. Yan, Z. Zheng, H. Li, and X. Liu, SID Symposium Digest **42,** 699 (2011).
11. X. Xia, X. Liu, H. Li, Z. Zheng, H. Wang, Y. Peng, and W. Shen, Opt. Express **21,** 11237 (2013).