Multi-viewpoint Image Array Virtual Viewpoint Rapid Generation Algorithm Based on Image Layering

Lu Jiang¹ and Yan Piao¹
¹Changchun University of Science and Technology, Changchun of Jilin, China

*E-mail: piaoyan@cust.edu.cn

Abstract. The use of multi-view image array combined with virtual viewpoint generation technology to record 3D scene information in large scenes has become one of the key technologies for the development of integrated imaging. This paper presents a virtual viewpoint rendering method based on image layering algorithm. Firstly, the depth information of reference viewpoint image is quickly obtained. During this process, SAD is chosen as the similarity measure function. Then layer the reference image and calculate the parallax based on the depth information. Through the relative distance between the virtual viewpoint and the reference viewpoint, the image layers are weighted and panned. Finally the virtual viewpoint image is rendered layer by layer according to the distance between the image layers and the viewer. This method avoids the disadvantages of the algorithm DIBR, such as high-precision requirements of depth map and complex mapping operations. Experiments show that, this algorithm can achieve the synthesis of virtual viewpoints in any position within 2×2 viewpoints range, and the rendering speed is also very impressive. The average result proved that this method can get satisfactory image quality. The average SSIM value of the results relative to real viewpoint images can reaches 0.9525, the PSNR value can reaches 38.353 and the image histogram similarity can reaches 93.77%.

1. Introduction
In the process of three-dimensional integrated imaging, traditional lens array recording systems have problems of low resolution and crosstalk that have long been difficult to solve [1]. At the same time, multi-view high-resolution image recording system (such as camera array) has a very broad application prospect in the acquisition of three-dimensional information of large scenes and the acquisition of high quality elemental images [2] because of independent non-interfering imaging units. Therefore, the combination of sparse scene viewpoint information and virtual viewpoint generation technology to achieve multi-view stereoscopic display has become the focus of the integrated imaging research. Currently, the mainstream algorithm for generating the virtual viewpoint image rendering technology is DIBR (Depth-Image-Based Rendering) [3]. However, its 3D mapping operation is very complicated and requires a large amount of computation, and the generating image quality depends on the depth image accuracy.

Aiming at the existing problems of virtual viewpoint technology, this paper presents a method which generate virtual viewpoint of camera array fast based on image layering. This method does not require the input of depth data and camera array parameters. Experiments show that this method can achieve the rendering of viewpoints anywhere within 2×2 viewpoint range.
2. Three-dimensional information acquisition based on multi-view image array

With the continuous development of three-dimensional display technology, the lens array recording system has been unable to meet the demand of obtaining large-area three-dimensional information due to its low resolution and crosstalk problems[4]. Therefore, people began to research the multi-view image array recording system to achieve the recording of the three-dimensional information recording.

As shown in the figure 1, camera array after the virtual viewpoint synthesized and display lens array has the same dimension $n \times m$. The focal length of the neutron camera in the camera array is shown as $f_1$ and the interval is $a_1$. The focal length of the lens element in the lens array is shown as $f_2$ and the interval is $a_2$. The center coordinates of the sub-camera 0 and the sub-camera n in the same column are respectively set as $(0,0,0)$, $(na_1,0,0)$. The center coordinates of the lens element m and the lens element n in the same columnin of the display lens array are respectively set as $(0,0,0)$, $(na_2,0,0)$. Take any two points on three-dimensional objects $P$ and $Q$, they can be imaged as pixel points $P_1$ and $Q_1$ on the elemental image array through the camera array. And then through the display lens array integration, we can get two points in the space of the integrated three-dimensional phase points $P_3$ and $Q_3$.

![Figure 1](image)

(a) Recording process  (b) Display process

Figure 1. Any two points on a three-dimensional object recording and display process

In practical application, the parameters of recording camera and display side lens are often difficult to match[5]. One-time recording of object three-dimensional information can not be applied to different display lens array with different parameters. If we use a sparse camera array to capture reference viewpoint information, the problem will be solved effectively.

3. Virtual viewpoint synthesis algorithm

3.1. Synthetic structure of two-dimensional viewpoint

The contribution of the two reference views to the virtual viewpoint synthesis effect is very limited, the background information occluded in the scene is difficult to complete. Therefore, this method of virtual viewpoint synthesis is based on multi-viewpoint, which can improve the virtual viewpoint information and the rendering image quality. Reference point of view images used in this article and their relative position shown as figure 2(a), the size of array is $4 \times 4$, the adjacent viewpoints have the same horizontal and vertical spacing, image resolution of $320 \times 240$.

Create a virtual view synthesis coordinate system with $0,1,4,5$ reference view as a plane(solid line selected area shows in figure 2(b), the viewpoints and their coordinate positions as shown in figure 2(c).

3.2. Obtain the depth information

Acquire scene depth based on 0 and 1 viewpoint images, this method requires less depth of image accuracy. Therefore, the SAD similarity measure is adopted as the stereo matching, which has the advantage of high speed[6]. The proceed shows as follows:
3.3. Two-dimensional image layering and disparity calculation

The obtained depth map is set as $I_d(x, y)$, the depth layer index value $Sign(x, y)$ corresponding to each pixel in the depth map is calculated according to the set number of layers $N$ ($N$ is a positive integer greater than 1). The formula is as follows:

$$Sign(x, y) = \lfloor I_d(x, y)/(256/N + 0.5) \rfloor$$

(1)

Where $\lfloor a \rfloor$ represents the largest integer operation less than or equal to $a$. Obtain the layered depth map $I_d^S(x, y)$ of the two-dimensional image according to the depth layer flag $Sign(x, y)$. The index of the stratified depth map $I_d^S(x, y)$ is $sign$, which is smaller the value, the closer the depth layer is to the viewer. When $Sign(x, y)$ equals to $sign$, $I_d^S(x, y)$ equals to $I_d(x, y)$, otherwise equals to 255.

Select a depth layer in the layered depth image as the focal layer of the camera array. Determine the foreground layer and background layer according to the focus layer. If the index value $sign > focus$, set it as the background layer; If the index value $sign < focus$, set it as the background layer. The reference viewpoint images are similarly stratified. Calculate the layered image $I_c^S(x, y)$ based on the layered depth image, at the same time record each layer images assignment flag $F_{c}^S(x, y)$. When $Sign(x, y)$ equals to $sign$, $I_c^S(x, y)$ equals to $I_d^S(x, y)$, $F_{c}^S(x, y)$ equals to 1. Otherwise, $I_c^S(x, y)$ equals to 255 and $F_{c}^S(x, y)$ equals to 0.

Let the maximum disparity value between the reference viewpoint foreground layer and its adjacent viewpoint fore-ground layer be $D_M$ ($D_M$ is an integer greater than or equal to one), calculate the maximum disparity value $R_M$ between the foreground and background layers:

$$R_M = (D_M / focus) \cdot (N-1)$$

(2)

Equation (6) shows how to compute the corresponding dis-parity value of each depth layer combine the distance of each layer to the focal layer

$$Dis^{sign} = -| (focus - sign) \times (R_M / (N-1)) + 0.5 |$$

(3)

Select the number of layers $N = 5$, the maximum disparity value $D_M = 30$, get layered result as shown in figure 3.

3.4. Virtual viewpoint rendering

Assuming that the size of virtual viewpoints in the virtual parallel multi-view array is $M \times M$ ($M$ is an integer equal to or greater than 2) and the zero reference viewpoint is the central viewpoint of the multi-view array. For the virtual viewpoint $i$ with coordinates $(a, b, 0)$, its index value of $sub_i$ can be count by the formula (7).
\[ \text{sub}_i = (M - b)M - \left| \frac{M}{2} - a \right| \quad i = 0, 1, 2, \ldots, M^2 \]  
(4)

The relative distance from the virtual viewpoint \((a, b, 0)\) position to the reference viewpoint position \((x, y, 0)\) is

\[ D_{mb} = \sqrt{(x-a)^2 + (y-b)^2} \]  
(5)

Where \(j\) denotes the position of the reference viewpoint, \(j = 0, 1, \ldots, 16\).

According to each layer of images corresponding to different disparity values, render the virtual viewpoint image based on weighted translation algorithm. Suppose we want to generate the image of the \(i\) position in \(M\) virtual viewpoints, we should perform a weighted panning operation on the image layer and the assignment flag according to the relative distance between the virtual viewpoint position and the reference viewpoint position and the disparity corresponding to each image layer. The equations are as follows

\[
I^i_h(x, y) = \begin{cases} 0 & (x + \text{Dis}^\text{vir} \bullet \text{Dis}^\text{vir} \bullet D_{mb,x} < 0 \text{ or } (x + \text{Dis}^\text{vir} \bullet D_{mb,y} \geq H) \\ I^i_c(x + \text{Dis}^\text{vir} \bullet \left| x - a \right|, y + \text{Dis}^\text{vir} \bullet \left| y - b \right|) & 0 \leq (x + \text{Dis}^\text{vir} \bullet D_{mb,x}) \leq H \\
\end{cases}
\]  
(6)

\[
F^j_h(x, y) = \begin{cases} 0 & (x + \text{Dis}^\text{vir} \bullet D_{mb,x} < 0 \text{ or } (x + \text{Dis}^\text{vir} \bullet D_{mb,y} \geq H) \\ F^j_c(x + \text{Dis}^\text{vir} \bullet \left| x - a \right|, y + \text{Dis}^\text{vir} \bullet \left| y - b \right|) & 0 \leq (x + \text{Dis}^\text{vir} \bullet D_{mb,x}) \leq H \\
\end{cases}
\]  
(7)

Where, \(H\) represents the image horizontal resolution. Initialize the virtual viewpoint’s assignment flag with position \(i F^i_{\text{vir}}(x, y) = 0\).

The translated images \(I^i_h(x, y)\) are superimposed layer by layer follow the order of putting the foreground in the background, that is, the order of \(\text{sign}\) from big to small. For Each time superimposed image layer, update the image assignment flag \(F^i_{\text{vir}}(x, y)\). After the overlay operation, a virtual viewpoint image with position \(i\) is obtained, the formula is as follows

\[
F^i_{\text{vir}}(x, y) = \begin{cases} 0 & F^i_h(x, y) = 0 \text{ and } F^j_h(x, y) = 0 \\ \text{sign} + 1 & F^i_h(x, y) = 1 \\
\text{sign} + 1 + N & F^i_{\text{vir}}(x, y) = 0 \text{ and } F^j_h(x, y) = 2 \\
\end{cases}
\]  
(8)

\[
I^i_{\text{vir}}(x, y) = \begin{cases} 0 & F^i_{\text{vir}}(x, y) = 0 \\ F^i_{\text{vir}}(x, y) - 1 & F^i_{\text{vir}}(x, y) > 0 \text{ and } F^i_{\text{vir}}(x, y) - N \leq 0 \\
F^i_{\text{vir}}(x, y) - N & F^i_{\text{vir}}(x, y) - N > 0 \\
\end{cases}
\]  
(9)

This rendering order, effectively ensure that the foreground object and the background object between the correct occlusion relationship. Perform the above operations on the \(M\) virtual viewpoints in turn, we can get \(M \times M\) virtual view of the image. The composite view at \((0,0,0)\) position of coordinate system shows in figure 4 below.

![Figure 3](image1.png)  
**Figure 3.** Layered schematic of depth images  

![Figure 4](image2.png)  
**Figure 4.** Virtual viewpoint image of (0,0,0)

### 4. Experimental results

For the coordinates in figure 2(c), the synthesized view point image at some position is shown in the figure 5 using the above method.
In order to verify the effect of the synthesized viewpoint and the true viewpoint, in the same way, viewpoint 5's image is rendered as a virtual viewpoint using reference view 0, 2, 8, 10 (as shown in the dotted line area in figure 2(b), the experimental results shown in figure 6.

In the subjective evaluation, the virtual viewpoint rendered image is already very close to the true viewpoint image, basically retained the complete image information. Only a slight blur appears in the foreground plaster and the lamp. In the objective evaluation, firstly, use color histograms as measures for image similarity comparison. The method of this paper is used to synthesize viewpoint 5, 6, 9 and 10. The virtual viewpoint images are shown in figure 7, and the data results shown in figure 8.

As can be seen from the figure 9, virtual point of view and the true point of view have a high degree of similarity and the mean of similarity is 93.77%.
Then, PSNR (peak signal to noise ratio)[7], SSIM (structural similarity index)[8] and $t$ (time) are used to objectively compare the quality of generated virtual viewpoint image, the unit of $t$ is s. Among them, the higher the PSNR value, the closer the SSIM value is to 1, the better the image quality is. The synthesis time for each viewpoint was also measured. Synthetic viewpoint images and the real viewpoint images are compared separately and the table below shows the data.

It can be seen in the table that all the numerical values are basically acceptable and the objective synthetic quality is satisfactory.

|      | PSNR   | SSIM   | $t$   |
|------|--------|--------|-------|
| 5    | 38.3369| 0.9533 | 6.7895|
| 6    | 38.3296| 0.9544 | 6.9209|
| 9    | 38.4082| 0.9511 | 7.0619|
| 10   | 38.3374| 0.9510 | 6.8302|
| average | 38.3530 | 0.9525 | 6.8998|

5. Conclusion

Based on the principle of integrated imaging, a virtual view rendering method based on image hierarchical algorithm is proposed in this paper. Firstly, the viewpoint image is layered according to the depth information of the scene, and then the weighted viewpoint translation of the layered image is used to obtain the virtual viewpoint image. The synthesis effect of different viewpoint is validated, experiments show that we can quickly get any position of virtual view images effectively, which laid a good foundation for the synthesis of elemental image array of lens array display system.

References
[1] JIAO Xiao-xue, ZHAO Xing, YANG Yong, FANG Zhi-liang, YUAN Xiao-cong 2012 Pick-up system for three-dimensional integral imaging with camera array Optics and J.Precision Engineering. 8. 1653-60
[2] YUAN Xiao-cong, XU Yu-ping, YANG Yong, ZHAO Xing, BU Jing 2011 Design parameters of elemental images formed by camera array for crosstalk reduction in integral imaging J.Optics and Precision Engineering. 9. 2050-56
[3] ZENG Yao-Xian 2012 A Dissertation Submitted to Shanghai Jiao Tong University for Master Degree of Engineering D.Shanghai Jiao Tong University
[4] Jiao Xiao-Xue, Zhang Lei, Sun Yu, Zhou Li-Qiu, Zhao Xing 2015 Three-dimensional reconstruction for a large scene of integral imaging based on color-position characteristics. J.Optoelectronics Letters. 4. 268-272
[5] Wang Yu, Piao Yan 2014 Computational Reconstruction for Integral Imaging with Sampled Elemental Images J.ACTA OPTICA SINICA. 5.70-75
[6] Yue Chen-ping, Meng Liya 2013 Implementation og the stereo matching based on SAD algorithm. J.Image Processing and Multimedia Technology. 6. 41-43
[7] PINSON M H, WOLFS 2004 A new standardized method for objectively measuring video quality J.IEEE Transactions on Broadcasting .50(3). 312-322
[8] WANG Z, BOVIK A C, SHEIKHHR 2004 Image quality assessment: from error visibility to structural similarity J.IEEE Transactions on Image Processing. 13(4). 600-612