An Efficient Method for All-in-focused Light Field Rendering

WEN Wei
Key Laboratory of Specialty Fiber Optics and Optical Access Networks, School of Communication and Information Engineering
Shanghai University Shanghai, China
3warrickwen@gmail.com

ZHANG Zhi Jiang
Key Laboratory of Specialty Fiber Optics and Optical Access Networks, School of Communication and Information Engineering
Shanghai University Shanghai, China
zjzhang@shu.edu.cn

YAO Si Cong
Key Laboratory of Specialty Fiber Optics and Optical Access Networks, School of Communication and Information Engineering
Shanghai University Shanghai, China
yaosicon86@shu.edu.cn

ZENG Dan
Key Laboratory of Specialty Fiber Optics and Optical Access Networks, School of Communication and Information Engineering
Shanghai University Shanghai, China
dzeng@shu.edu.cn

Abstract—Light field rendering (LFR) is an image-based method for generating novel views from a set of camera images. When sample of camera array is sparse, conventional light field rendering would cause aliasing artifacts. We present a straightforward layer-based method that significantly removes the aliasing artifacts with no prior knowledge of the scene geometry. First, a sparse disparity map is generated to detect the depth layers of the scene. Then, for synthesized each pixel, using the color consistency constraint to detect the optimum focal plane from the layer depths, and rendering such pixel with selected focal plane by LFR method. Finally, a trick method is used to reduce flickering artifacts caused by the color consistency constraint. Experimental results show the effectiveness of our approach.

Keywords-Image-based rendering; light field rendering; all-in-focus; depth layers; disparity

I. INTRODUCTION

In recent years, Image-based rendering (IBR) has been found widespread use because of the ease with which it produces photorealistic imagery. IBR methods do not require prior knowledge of the scene geometry. Instead, IBR requires a large number of reference images taken with densely arranged cameras. A formal analysis of the tradeoff between the sample rate and the amount of geometry is presented in [7].

Light field rendering (LFR) [1] is one simple algorithm of IBR methods, which could render novel views from a dense set of reference views. Each pixel value of input images is parameterized as a light-ray data, and stored in a light-ray database. We can synthesize a view of a given viewpoint from the database by picking up the data of such light-rays those pass through the view point. In most practical situation, however, since the sample of camera array is sparse, the database contains only a finite sampling of the rays. In this “under sample” case, when rendering a large depth scene, single focal plane in LFR will cause obvious aliasing artifacts.

In order to suppress aliasing artifacts, Levoy and Hanrahan [1] eliminated aliasing artifacts with proper pre-filter, however, this pre-filtering results in the degradation of the rendering quality. J. Stewart [3] presented a new reconstruction filter that suppresses artifacts by cutting the high frequency components of out-of-focus region, and preserves textures at in-focus region by wide-aperture reconstruction [2]. But these kinds of methods can not render an all in-focus novel view.

Another solution is to use multiple focal planes. Takeshi Naemura [5] performed stereo reconstruction to generate a dense disparity map, and rendering each synthesized pixel with different depth of focal plane. Unfortunately, the state of automatic stereo algorithms is inadequate for producing sufficiently accurate depth information for realistic rendering. Minglun Gong [7] presented a color-matching based interpolation, which searches for a possible physical point along the testing ray using the color consistency constraint in nearby reference images. But this algorithm leads to a huge computational cost and noise. Suppose that having the prior knowledge of the range of scene depth, Tauchii. Y [4] assumed a layered depth model in the object space to equally divide the scene and estimate the depth of focal-plane by the color consistency constraint for each target light ray. In practice, it is hard to gain the range of scene depth. Moreover, equally dividing will result in extra noise.

Our approach follows the layered depth model and the color consistency constraint. Geometry (depth information) estimation has been used to compensate for “under sample” of camera array. The method overcomes the disadvantages in [5, 7, 4]. Given a novel viewpoint, two calibrated cameras are selected to generate a sparse disparity by stereo algorithm to detect the depth layers of scene. Then, for each
synthesized pixel, using the color consistency constraint to

detect the optimum focal plane from the layer depths, and

rendering such pixel with selected focal plane. Finally, in

order to reduce flickering artifacts caused by camera noise

and other factors, a trick method applied to smooth the each

layer.

The contributions of this paper are outlined as follow:

- Without knowing the prior knowledge of the scene

  geometry, our algorithm totally suppresses the

  aliasing artifacts caused by "under sample".

- In contrast to Tauchini. Y’s equally dividing the

  scene, our method divide the scene by a robust

  sparse disparity algorithm. Due to the characteristic

  of our method, it overcomes [5, 7, 4]'s disadvantage

  of huge computational cost and noise.

- A simple smooth function has been applied to reduce

  flickering artifacts caused by camera noise and other

  factors.

II. Rendering Algorithm

We assume that multi-calibrated cameras lie on a plane

and are arranged on a 2D grid. There is no prior knowledge

of the scene geometry. Fig.1 represents the algorithm flow of

our method. Firstly, a robust sparse disparity algorithm is

presented to detect the depth layers of scene. Then optimal

depth layer is assigned to each synthesized pixel as depth of

focal-plane. Finally, by apply a simple smooth function to

reduce flickering artifacts caused by the camera noise and

other factors, and an all-in-focus view is reconstructed where

aliasing artifacts are significantly suppressed.

II.1 Layered algorithm with sparse disparity

To render each pixel on the synthesized image by optimal

focal-plane, it’s needed to know the depth layers of the scene.

In the practical application, dynamic scenes will be treated

and detecting depth layers at each frame. Based on the

mentioned consideration, layered method should be satisfied

with high robust and low computational cost. However, it is

hard for automatic stereo algorithms to get a sufficiently

accurate dense disparity map. Therefore, this paper

introduces a layered method detects scene layer by the

distribution function of sparse disparity.

Firstly, an efficient matching algorithm for sparse

disparity is presented, which is based on SAD operator [9]

and fix window searching method. In this paper, we set size

of search window to $7 \times 7$ and set a low SAD threshold thus

reducing the error caused by occlusion. The sparse disparity

map of “Tsukuba” is show as Fig.2 (left) (pixel value is
twelve times disparity value):

![Sparse disparity map and distribution function of “Tsukuba”](image)

As shown in Fig.2 (left), there are occlusions around

scene edge that lead to error, thus algorithm set it to zero.

Nevertheless, distribution function of this sparse disparity is

consistent with that of scene depth. It can be seen from Fig.2

(right) in the form of graph. From the graph, we learn that

most of scenes are mainly distributed in five layers and lots

of error layers are in that distribution.

Based on the above analysis, a novel de-noising method

is proposed, which introduce a ratio coefficient $\lambda$ that set

the error tolerance to eliminate error layers. $\lambda$ represent

the ratio of the width (height) of error area and the width (height)

of synthesized image, which is defined as follow:

$$\lambda = \frac{e_{\text{width}}}{W} \Rightarrow e_{\text{width}} = \lambda \cdot W$$

$$\lambda = \frac{e_{\text{height}}}{H} \Rightarrow e_{\text{height}} = \lambda \cdot H$$

(1)

$$S_{\text{area}} = e_{\text{width}} \cdot e_{\text{height}}$$

where $W$ (H) denotes the width (height) of synthesized

image, and $e_{\text{width}}$ ($e_{\text{height}}$) denotes the width (height) of error

area. $S_{\text{area}}$ represents the error tolerance. If the number of

pixel in certain layer is lower than $S_{\text{area}}$, it will be treated as

error layer. We use the following equation to eliminate the

error layer:

$$\text{plane}(n) = \begin{cases} 
  \text{false} & \text{if } \text{pixelNum}(n) < \lambda^2 \cdot \text{sumpixelNum} \\
  \text{true} & \text{otherwise}
\end{cases}$$

(2)

Where $\text{plane}(n)$ denotes n-th layer in sparse disparity

map, $\text{pixelNum}(n)$ denotes number of pixel in n-th layer,

$\text{sumpixelNum}$ represent the sum number of pixel in sparse

disparity except zero pixels.

The layered algorithm above use distribution function of

sparse disparity to divide scene into layers, and a novel de-

noising method is proposed to eliminate most error layers,

but there are still small number of error layers. Fortunately, a

trick method will be applied to reduce flickering artifacts

caued by error layers. The scene of “Tsukuba” is divided

into 7 layers.
II.2 Mutil-focal plane LFR with the restriction of color consistency

By layered algorithm above, we divide scene into a set of depth layers. In this part, we assign optimal depth layer from depth layers to each synthesized pixel by the color consistency constraint, and render it by conventional light field rendering method.

As with the parameterization in [2], in this paper we index the camera position using \((s,t)\) and use \((u,v)\) as the pixel position on the imaging plane at each camera. The focal plane is defined as a depth plane that we assume in the scene when rendering a novel view by LFR. For simplicity, consider a two dimensional version of our light ray parameterization where parameters \(t\) and \(v\) are fixed, as show in Fig.3:

![Figure 3. Light field parameterizations and rendering method](image)

Let \(r(x)\) be a novel ray that is rendered with the virtual camera \(C_x\) at view position \((s_z,t_z)\), in which \(z_r\) denotes the distance from \(S\) axis. \(Focalplane_1\), \(Focalplane_2\), \(Focalplane_3\) are the focal plane which arranged with depth layers of scene. The novel ray intersects with three focal-planes at \(P_i^n\) \((n=0,1,2)\), in which \(P_i\) is the intersection of scene. Before introducing the color consistency constraint, we need to make the following two assumptions:

- Any point in the scene that is visible from the novel view point is also visible in the nearby four camera.
- The projections of the same physical point in the scene should have a higher level of color similarity than the projections from different physical points.

If the above assumptions hold, we know that the re-projection of intersection of scene should have the highest level of similarity in color. Hence, what we need to do is to project those intersections to nearby reference camera. The point, whose re-projection give the smallest dissimilarity, will be the intersection require. As shown in Fig.3, \(p(s_0,u_0), p(s_1,u_1)\) represent the pixel value obtained from the re-projection of \(P_1\). While, \(p(s_2,u_2^-), p(s_3,u_3^-)\) represent the re-projection of \(P_2, P_3\) respectively. The evaluation functions of color consistency can be calculated as follow, and the smallest dissimilarity leads to lowest value of \(Consistence(z)\):

\[
Consistence(z) = \frac{|p(s_1,u_1) - p(s_2,u_2^-)|}{2}
\]  

Each pixel in synthesized image will be assigned an optimal depth plane by the color consistency constraint above. In Fig.3, \(Focalplane_2\) is assigned to render novel ray \(r(x)\) at the given pixel \(x\) is rendered using pixels corresponding to \(P_2\), and is given by:

\[
r(x) = w_i l_i(s_i,u_i) + w_j l_j(s_{i+1},u_{i+1})
\]

Where \(w_i = \frac{s_i - s_{i+1}}{s_{i+1} - s_i}\), \(w_{i+1} = \frac{s_{i+1} - s_i}{s_{i+1} - s_i}\)

Fig.4 shows the experimental results for our rendering method. Although aliasing artifacts have been almost suppressed, there are extra flickering artifacts on synthesized image due to error layers and camera noises. As shown in Fig.4, Noise is found in “book” and smooth shadow disappears from “lamp”.

![Figure 4. Experimental results for multi-focal plane](image)

The next section, we will analyze the origin of the noise. Meanwhile, a smooth function will be introduced to eliminate flickering artifacts, and experimental results will show that method can render photo-realistic views.

II.3 Suppress the flickering artifacts

As the camera noise, the color of scene and other factors, color consistency constrains will introduce a serious flickering artifacts. As shown in Fig.5, we only analyze the condition of two focal-planes, and the same can be extended to multi-focal-plane. When rendering a novel ray \(r(x)\), color
consistency constraint will be used to detect the optimal focal plane from those two planes. \( p(u_0, s_0) \) and \( p(u_1, s_1) \) are the re-projection of \( P_i \), and they all represent the color of \( P_i \). \( p(u_0', s_0') \) and \( p(u_1', s_1') \) are the re-projection of \( P_j \), however, they represent the color of \( e_0 \) and \( e_1 \). In Fig.5, due to the color of scene or camera noise, the re-projections of \( P_j \) give the smaller dissimilarity, and lead to the false judgment:

\[
\text{Consistence}(Z_{focalplane2}) < \text{Consistence}(Z_{focalplane1})
\]

Therefore, \textit{focalplane2} was assigned to novel ray \( r(x) \) as optimal focal-plane, and result is show at left-down corner of Fig.5.

![Schematic diagram of flickering artifacts](image)

**Figure 5. Schematic diagram of flickering artifacts**

From the above discussion, the flickering artifacts arising from the color of scene or camera noise, and error layers will result in an increased possibility of flickering artifacts. \cite{4} uses the changes in two adjacent frames to eliminate flickering noise, whereas we suppress it in one frame with smooth focal-plane depth map. As shown in Fig.6,

![focal-plane depth map](image)

**Figure 6. focal-plane depth map**

The value of each pixel represents the depth map of optimal plane depth. Sudden changes in focal-plane depth lead to disappearance of smooth shadow on “lamp”. Suppose that the depth of the scene at a particular object changes continuously, so focal-plane depth map should also changes continuously at a particular object. Based on the above considerations, sudden changes of pixel value represent the location of flickering artifacts. Here, we present a smooth function to locate flickering artifacts and re-render it with proper focal-plane:

\[
\hat{P}(x, y) = \begin{cases} 
R \text{render}[x, y, \text{med}[\text{fmap}(x, y)]] & \text{if } \text{fmap}(x, y) = \text{med}[\text{fmap}(x, y)]; \\
\hat{P}(x, y) & \text{otherwise}; 
\end{cases} 
\]

Where, \( \hat{P}(x, y) \) and \( P(x, y) \) denote the synthesized pixel before and after suppressing flickering artifacts respectively. \( \text{fmap}(x, y) \) denotes the pixel in focal-plane depth map. \( \text{med}[\text{fmap}(x, y)] \) represents taking middle value at the adjacent area of position \((x, y)\) in focal-plane depth map. The smooth function above shows that if the value of position \((x, y)\) in focal-plane depth map is not equal to its middle value of adjacent area, it will be regard as position of flickering artifacts. Then, re-render this synthesized pixel with correct focal-plane, which is reset by its middle value of adjacent area. Fig.7 and Fig.8 show the experimental result of focal-plane depth map and final synthesized view.

![smooth focal-plane depth map](image)

**Figure 7. smooth focalplane depth map**

![final all-in-focus synthesized view](image)

**Figure 8. final all-in-focus synthesized view**
III. EXPERIMENTAL RESULTS

We use 25 reference images captured from a 5×5 camera array, which are provided from “The Multiview Image DataBase,” courtesy of the University of Tsukuba, Japan. Image resolution is 320×240 (down sampling from original image). The horizontal and vertical distance between is 20mm. Fig.8 shows the experimental result of our rendering method, which suppressed aliasing artifacts. As shown in Fig.9, the novel view in left part was reconstructed by [4]. Due to equally dividing of scene, a large number of error layers were introduced. However, we divide depth of scene by the distribution function of sparse disparity. The comparison result shows that our approach provides more details in rendering high frequency information than [4].

IV. CONCLUSIONS

We proposed a novel image-based rendering method that uses inputs from an array of cameras and synthesizes free-viewpoint images by using a set of depth layers. Our approach divide scene into a set of depth layers and optimal depth layer is assigned to each pixel on the synthesized image by color consistency constraint. A smooth function is introduced to eliminate the flickering artifacts caused by camera noise. Finally, our approach produce photo-realistic image at a low computation cost.

Our algorithm combines the merits of previous rendering methods. In addition, the advantage of the proposed method is that all-in-focus views can be rendered with no prior knowledge of the scene geometry in “under sample” condition.

ACKNOWLEDGMENT

This work is supported by State Key Program of National Natural Science Foundation of China (No. 60772124) and by Shanghai Leading Academic Discipline Project and STCSM (S30108 and 08DZ2231100). The authors are with the Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, 200072 Shanghai.

REFERENCES

Figure 10. the experimental result of “Tsukuba”, (a) Tauchi’s method, (b) our method