Triangulated Views Selection for Image-based Walkthrough

Chang Ok Yun, Jung Hoon Kim, Tae Soo Yun and Dong Hoon Lee

***Abstract*— This paper presents an adaptive sampling method for image-based walkthrough. Our goal is to select minimal sets from the initially dense sampled data set, while guaranteeing a visual correct view from any position in any direction in walkthrough space. With regard to image based representation, we adopt image based rendering by warping method. Especially we utilize three reference images for warping, that means the goal of the paper is to find the optimized sample set with respect to the three reference images. In this paper, we regard to the problem as approximating the mesh model, which is constructed from initial dense data and addresses an optimization scheme through global search with an approximated error measurement.**

***Index Terms* — Image-based walkthrough, Adaptive sampling method, Image-based rendering by warping, Optimization, Mesh simplication**

I. INTRODUCTION

Image-based rendering (IBR) generates novel views from a set of input images instead of 3D models. Among the many IBR approaches, one promising IBR approach enhances the images with per pixel depth. This allows warping the samples from the reference image to the desired image. Generally the image-based approach using depth is called as image-based rendering by warping (IBRW) [7]. However, in IBRW, simply warping the samples does not guarantee high-quality results because one must reconstruct the final image from the warped samples. To solve this problem, most of all, good reconstruction algorithms such as efficient warping, splatting etc. are needed [6, 8]. But the fundamental and important problem of properly sampling has remained largely unanswered.

The sampling is a very difficult problem since the sampling rate will be determined by the scene geometry, the texture on the scene surface, the reflection property of the scene surface, the specific IBR representation we take, the capturing and the rendering camera's resolution, etc [12]. Over-sampling was widely adopted in the early stages [5, 11]. Generally, to reduce the huge amount of data due to over-sampling, many compression techniques are utilized [12] instead of dealing with the sampling problem. However, sampling is more of a fundamental problem to IBR.

In this paper, we basically deal with the sampling problem, where the goal is to select minimal sets while guaranteeing a visually correct view from any position in any direction. The object of this research is to devise a method for the efficient re-sampling of initial pre-sampled data and for the efficient management of the data for interactive walkthrough.

With regard to image based representation, we adopt image based rendering by the warping method. Especially we utilize three reference images for warping, that means the goal of the paper is to find the optimized sample set with respect to the three reference images. Therefore, the particular problem we would like to address in this paper is expressed as follows:

We are given dense sampled images, each with depth and calibration information at a given position. We are also given a reconstruction error estimator, which estimates the quality of the reconstructed image from three reference images. From the set of initial samples, we have to determine a set of optimized three reference images, while keeping the error within the requirement.

II. BACKGROUND AND PREVIOUS WORK

Depending on how the capturing cameras are placed, IBR sampling can be classified into two categories: uniform sampling and non-uniform sampling. In uniform sampling, the cameras are positioned evenly on a capture configuration which is usually a surface or a line. The light field [5] and the concentric mosaics [11] are the representative examples. In the case of uniform sampling, the main research topic is to find the minimum sampling rate or largest spacing between cameras such that one can achieve perfect reconstruction quality on the navigation space. The goal of non-uniform sampling analysis is also to find the minimum number of cameras while rendering the highest quality scene. But in this case, arranging camera position is another important problem.

In practice, objects in the scene have varying surface properties. For instance, if a scene has non-Lambertian surface or occluded regions, more samples are needed. In general, a real world scene is composed of Lambertian and non-Lambertian surfaces. The Lambertian surface may need a low sampling rate, while the non-Lambertian surface may

Chang Ok Yun is with Department of Visual Content, Graduate School of Design&IT at Dongseo University, San 69-1 Churye-2- Dong, Sasang-Gu, Busan, Korea(E-mail:coyun@hanmail.net).

Jung Hoon Kim is with Department of Visual Content, Graduate School at Dongseo University, San 69-1 Churye-2- Dong, Sasang-Gu, Busan, Korea(E-mail:melc81@gmail.com).

Tae Soo Yun is with Department of Visual Content, Graduate School of Design&IT at Dongseo University,San 69-1 Churye-2- Dong, Sasang-Gu, Busan, Korea(E-mail:tsyun@dongseo.ac.kr).

Dong Hoon Lee is with Department of Visual Content, Graduate School of Design&IT at Dongseo University,San 69-1 Churye-2- Dong, Sasang-Gu, Busan, Korea(corresponding author to provide TEL:+82-51-320-2083; FAX:+82-51-322-2673; E-mail:dhl@dongseo.ac.kr).

need a high sampling rate. Uniformly sampling the scene without concerning about the regional surface property may easily cause over-sampling of the Lambertian surface or under-sampling of the non-Lambertian surface [12]. It is therefore natural to consider non-uniform sampling. One thing to notice is, in uniform sampling, since the sampling is periodic, we only need to tell how many images/light rays are needed for perfect reconstruction of the scene; in non-uniform sampling, however, we need to answer not only how many images are needed but also where to place these cameras. In this paper, the non-uniform sampling approach is adopted.

Fleishman et al. [2] proposed an automatic camera placement algorithm for IBR. They assumed a mesh model of the scene is known. The goal is to place the cameras optimally so that the captured images can form the best texture map for the mesh model. They proposed an approximation solution for the problem by testing a large set of camera positions and selecting the ones with higher gain rank. Here the gain was defined based on the portion of the image that can be used for the texture map. However, this method cannot be extended to the sampling of real environment due to the assumption of the known mesh model and is only applicable to scenes with Lambertian surfaces.

Schirmacher et al. [10] proposed an adaptive acquisition scheme for a LightField setup. Assuming the scene geometry is known, they added cameras recursively by predicting the potential improvement in rendering quality when adding a certain view. They asserted a priori error estimator accounts for both visibility problems and illumination effects such as specular highlights used to some extent. However, this error estimator definitely has its limits unless a ground truth data is obtained.

Depending on the application scenario, solutions to the non-uniform sampling problem can be classified into two categories: incremental sampling and decremental sampling [12]. In incremental sampling, the samples are captured one by one incrementally. The stopping criterion is either the overall number of samples desired reached, or an error requirement that the sampling process must achieve. In decremental sampling, we assume there is already a dense set of samples available that fulfills the error requirement. This sample set might be too large, thus decremental sampling can be used to reduce the size, while keeping the error within the requirement.

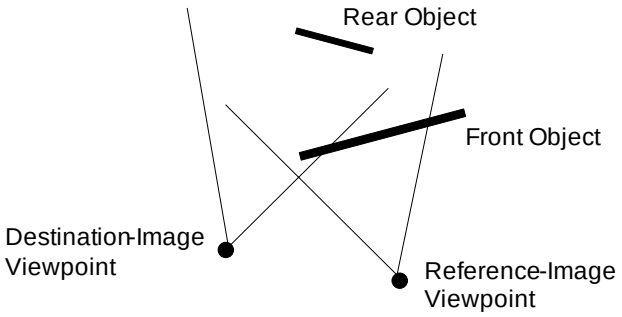


Fig. 1. A simple visibility example. The rear object is visible in the destination image, but occluded in the reference image [6]

Most previous image-based walkthrough systems which have the sampling strategy have adopted the incremental sampling approach since it was too difficult to get the initial dense samples. However, recent advances in acquisition and modeling technologies have resulted in large databases of real-world and synthetic environments. This is made possible by using an omni-directional camera capture system in which the camera is placed on a motorized cart together with a battery, computer, frame grabber and fast disk system to store the captured images on [1]. Therefore, in this paper, we choose the decremental sampling approach to determine optimal view positions.

III. SAMPLING FOR TRIANGULATED VIEWS

The sampling method in the decremental approach can be expressed as a view selection problem and the optimized views imply a kind of best view which is representative of the sampling space. We can consider a single reference image for warping. However, in most circumstances, a single reference image is insufficient to avoid unacceptable occlusion artifacts. Figure 1 shows such a case. The rear surface is not visible in the reference images, but should appear in the destination image. As a result, if we warp the reference image to the destination viewpoint, we will get a destination image that does not contain the rear surface. Figure 2 shows this phenomenon for a reference image and corresponding destination image.

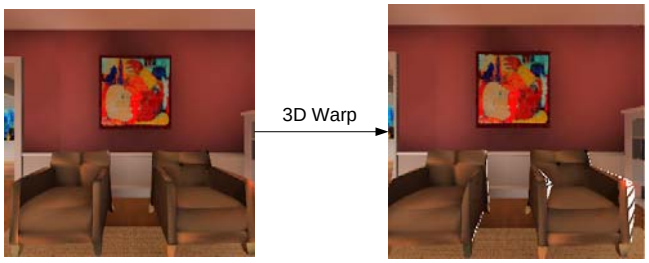


Fig. 2. The 3D warp can expose areas of the scene for which the reference frame has no information (shown here in white).

With respect to the region which involves the occlusion artifacts, the single reference image based best view selection chooses many samples to guarantee an adequate warping quality. Due to this reason, if we consider only a single reference image for best view selection, it is not sufficient to say that the selected set is fully optimized - although this can optimize the number of the samples in a fashion. In order to get acceptable quality of warped output, we need additional information about the scene. The post-rendering warping system [6] renders these additional images using view-points that are different from those used to render the first image.

A. Triangulation of Reference Views

There are several tradeoffs involved in choosing the number of reference images that are warped to produce each displayed frame. First, more reference images demand a high computational cost, because all of the reference images must be warped for every frame. Second, as we increase the

number of reference images, we also increase the lifetime of each reference image. The reason for this tradeoff is that the conventional rendering engine only produces one reference image at a time. We considered the cases using one or more reference images. We quickly ruled out the use of one reference image, because occlusion artifacts were too severe with only one reference image. After considering a greater-than-one image approach, we settled on the three-image approach. Using three images eliminates most visibility artifacts, while limiting both the cost of warping and the length of time that reference images remain in use.

Another reason is related to the working set management [1]. The selected sampled set is a kind of unorganized scattered data set. From the performance of the system perspective, the unorganized set affects other problems as follows:

First, it is too difficult to pick reference-frame viewpoints that lie on the user's path through space: A reference frame can be determined with a viewpoint near a previous viewer position. If we consider two reference images based warping, the second reference image is selected with a viewpoint near a 'future' view position. In most systems with an unorganized data set [1, 6], the future viewpoint is determined using a motion prediction algorithm. When the viewer passes this 'future' viewpoint, the system starts using a new 'future' reference image and discards the old 'past' reference image. However, generally, motion prediction is not perfect. As the prediction interval grows longer, the expected magnitude of the error in the prediction also grows. The characteristics of the prediction error will vary depending on how viewpoint motion is controlled.

Similarly, establishing cache mechanism for interactive walkthrough is also hard with the unorganized data set. The problem is related to pre-load and cache images from the disk for rendering novel views as a user navigates interactively through the IBR environment. In order to do this, we have to develop a time-critical algorithm that loads and caches data. The exhaustive search of the unorganized data set is not suitable for a time-critical algorithm.

Eventually, the above problems can be solved by organizing the data set. In a viewpoint picking and caching, the obvious problem is to find the nearest reference frame at a viewpoint. The organization of the scattered data in a metric manner is considered as follows:

Denote the Euclidean distance between two points p and q by dist(p, q) in R^2. Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of n distinct points in the plane; We can define the subdivision of the plane into n cells, with the property that a point q lies in the cell corresponding to a site p_i if and only if dist(q, p_i) < dist(q, p_j) for each $p_j \in P$ with $j \neq i$. A subdivided plane with the property is known as *Voronoi diagram*. From the Voronoi diagram of the sample data set, the nearest reference viewpoint can be estimated. As mentioned before, since a single reference image causes severe occlusion artifacts, other reference images are also needed. The dual of the Voronoi diagram - Delaunay triangulation, is suitable for this mechanism. By constructing the triangulated structure, the post-rendering warping can be accomplished effectively by fetching the three vertices (reference images) of a triangle.

B. Triangulated Views Selection

With respect to a single reference image based warping, the best view selection can be defined as finding a set of a single reference image which can warp the entire walkthrough space. However, the selection of the triangulated data set is totally different. Here one has to consider how to select three reference images which enable one to represent all positions in each triangle fully with the minimum triangles. Considering determination of a single optimized triangle, the number of cases is ${}_nC_3$, which means the time complexity is $O(n^3)$. Accordingly, the optimization about full space is a more difficult and complex problem. In addition, since the triangle has 3 vertices and the vertices hold some adjacent triangles, the selection of a vertex implies that every triangle that has the vertex selects the position as a best view position with respect to all the positions which the triangles contain.

To solve the problem, it can also be expressed in other ways: Given an initial dense triangle data set, it transforms the data set into the approximated data set with provable guarantees. We found a similar problem in computer graphics, which uses *mesh simplification*.

Among the previous approaches in mesh simplification, the edge decimation method shows very good simplification quality though the algorithm speed is slow. So, we adopt mesh decimation method for sampling.

C. Triangulated Views-Selection based on Edge Collapse

Our simplification algorithm is based on the iterative contraction of vertex pairs, which has generally been used in previous work. A pair contraction, which we will write as $(v_1, v_2) \rightarrow \bar{v}$, moves the vertices v_1 and v_2 to the new position $\bar{v}$, connects all their incident edges to v_1, and deletes the vertex v_2. Subsequently, any edges or faces which have become degenerate are removed. The effect of a contraction is small and highly localized. If (v_1, v_2) is an edge, then one or more faces will be removed (see Figure 3).

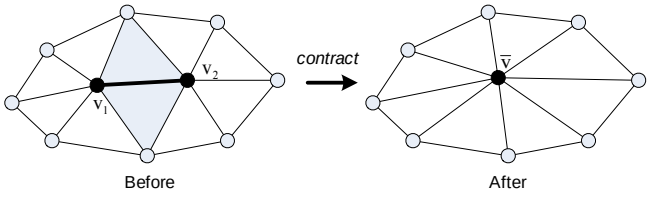


Fig. 3. Edge collapse. The highlighted edge is contracted into a single point. The shaded triangles become degenerate and removed during the contraction.

The next consideration is how to select the next edge to collapse. The trick to producing good optimized results is to select the edge that, when collapsed, will cause the smallest visual quality change to the entire warped position. The proposed collapse process can be divided into two steps:

a. compute all edge collapse cost,
b. perform the edge collapse operation

To define the cost, we attempt to characterize the error at

each edge. The error at an edge means the average error of the updated triangles which are jointed by the single point when the edge is contracted into a single point. The smaller error at an edge specifies the elimination of the edge has little effect on the visual quality of each warped position within the triangles. In this case, the collapse operator does not affect the original triangle structures. By performing the edge collapse hypothetically, we can derive the cost error. The overall procedure of computing edge collapse cost is described in Figure 4.

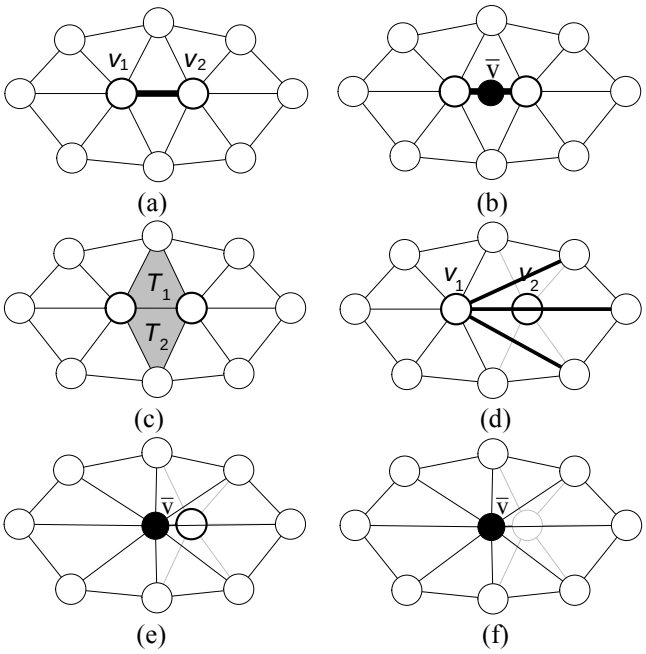


Fig. 4. The procedure of computing edge collapse cost. (a) select a valid pair, (b) compute the contraction target $\bar{\mathrm{v}}$ for each valid pair $(\mathrm{v}_1, \mathrm{v}_2)$, (c) remove any triangles (T_1 and T_2) that have both v_1 and v_2 as vertices hypothetically, (d) update the remaining triangles that use v_1 as a vertex to use v_2 instead hypothetically, (e) change the value of $\bar{\mathrm{v}}$ into the value of v_1 hypothetically, (f) remove vertex v_2 hypothetically and compute the error of the updated triangles which are jointed by the contraction target $\bar{\mathrm{v}}$. The error becomes the cost of edge $(\mathrm{v}_1, \mathrm{v}_2)$. This procedure is repeated until all valid pairs are visited.

The initial computed cost is placed in a heap, which places the minimum cost pair at the top. The edge collapse operation iteratively removes the pair $(\mathrm{v}_1, \mathrm{v}_2)$ of least cost from the heap, contracts this pair and updates the costs of all valid pairs involving v_1.

In order to perform the contraction $(\mathrm{v}_1, \mathrm{v}_2) \rightarrow \bar{\mathrm{v}}$, we must choose a position for $\bar{\mathrm{v}}$. The candidates for the new position $\bar{\mathrm{v}}$ are all vertices connected with the vertices v_1 and v_2. It would be nice to find a position for $\bar{\mathrm{v}}$ which minimizes error function $\Delta(\bar{\mathrm{v}})$. Eventually the determination of new position $\bar{\mathrm{v}}$ depends on which candidate position has minimum average error. However, since the computational cost for finding new position is too high, we use a simple scheme which selects either v_1, v_2 or $(\mathrm{v}_1 + \mathrm{v}_2)/2$ depending on which one of these produces the lowest value of $\Delta(\bar{\mathrm{v}})$.

D. Estimating the Reconstruction Error

To estimate the cost of each edge, the average error of the updated triangles is utilized. The error of each triangle can be estimated by the average error of the warped images with respect to the overall position in a triangle. However, since the cost of computing overall position in each triangle is too high, we use the centroid (center of gravity) of the triangle as a representative destination frame to approximate error value.

Every pixel i from every selected reference frame is warped to its new location in the destination image. For every destination pixel j with coordinates (u_j, v_j), let the source map $M_j=\{i_0, i_1, i_2\ldots\}$ be the set of pixels from different source images which map to (u_j, v_j). Among the warped pixels, the front most pixel $i_{\max}$ with the maximal Z value $z(i_{\max})$ from among M_j must be determined. Subsequently, we construct the blend map $M_j^{'}$ which is the subset containing all pixels at a depth within a ε-interval around $z_{\max}$:

$$M_j^{'} = \{ i \in M_j \mid\mid z(i) - z(i_{\max}) \mid < \varepsilon \}. \tag{1}$$

The final color of the pixel is determined by blending from all pixels in $M^{'}$ using a weighted sum:

$$d_j^{(B)} = \frac{1}{\sum w_k} \sum_{k=0}^{|M_j^{'}|} w_k Color(i_k). \tag{2}$$

The weights account for the importance of each source image with respect to the final image and can be chosen as the inverse distance between the destination view point and the view point associated with the source pixel.

To evaluate the performance of the 3D warping algorithm, we need a quantitative way to estimate the quality of the warped images. Two general approaches to this are to compute error statistics with respect to some ground truth data and to evaluate the synthetic images obtained by warping the reference by the computed disparity map. In our case, since we already sampled all images at every sample position, the evaluation of the performance is achieved by ground truth data using the following two quality measures:

a. RMS (root-mean-squared) error between the computed warped image $d_j^{(B)}$ and the sampled ground truth image $d_j^{(T)}$,

$$R = \left(\frac{1}{N} \sum_{j=0}^{N-1} \left| d_j^{(B)} - d_j^{(T)} \right|^2 \right)^{\frac{1}{2}}, \tag{3}$$

where N is the total number of pixels in the destination image.

b. Percentage of bad matching pixels,

$$B = \frac{1}{N}\sum_{j=0}^{N-1}\left(\left|d_j^{(B)} - d_j^{(T)}\right| > \delta_d\right), \quad (4)$$

where δ_d is a disparity error tolerance.

Accordingly, the total reconstruction error can be expressed as follows:

$$E = \alpha R + \beta B. \quad (5)$$

We select the weight of RMS (α) as 10.0 and the weight of percentage of bad matching pixels (β) is 0.1. We think the bad matching pixels have a great influence on the quality of scene and adjusted the parameters until we found an adequate weighting value. In addition, the disparity error tolerance δ_d sets to 15.0.

IV. EXPERIMENTAL RESULTS

The proposed algorithm was implemented on a Pentium IV PC with a 1.7GHz CPU and 521Mbyte memory. We captured a synthetic environment using 3D Studio Max. We captured three views at a location with 90° FOV and acquired a dense Sea of Images through a large environment, which yields an image every couple of inches (see Figure 5)[4].

Four color and depth map images at the same camera position were stitched into a panoramic image. At that time, we captured the rays with regular angular resolution. The amount of the test data is 450 images which are 15 stitched images with respect to 30 paths. The sample data in Figure 6 shows one of the synthesized panoramic images.

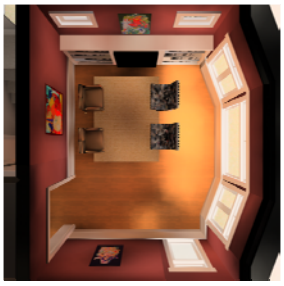
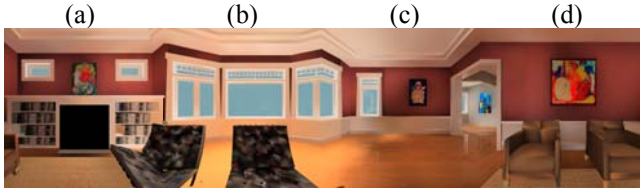


Fig. 5. A scene environment which we utilize for experiments. We captured 15 stitched images with respect to 30 paths.

(a) (b) (c) (d)



(e)

Fig. 6. Sample image of input panorama sequence. We have captured four views at the same location as illustrated in (a), (b), (c) and (d) respectively. The (e) is a result of stitching with them.

In order to verify the effectiveness and robustness of the edge collapse scheme for triangle views selection, several experiments were performed. The performances of the experiment results are evaluated by the reconstruction error as shown in section 3.4. To compare the result of the decimation, we also estimate the errors on other two test data. First, we triangulate initial dense data sample (a) and estimate the reconstruction error. In this case, it is needless to say that the reconstruction error of the triangle data of the initial sample is superior to the result of the proposed method. This experiment plays a role in determining the criterion on the best reconstruction quality. Second, we compared the proposed method (c) with the uniform sampling method (b) (when given the same number of decimated samples). For the uniform sampling, multiple path-based capture configuration is utilized[4]. Table 1 shows the experimental results on the triangle views selection based on edge collapse.

Table 1 Experimental results on the best triangle views selection algorithm.

Test sample	# of sample	# of triangle	mean error	std. deviation
Initial sample (a)	450	836	1.360674	0.056641
Uniform sample (b)	21	24	2.461836	0.618503
Edge collapse (c)	21	30	1.769296	0.278682

The test scene is captured with reference depth images according to the multiple path based capture configuration. The scene is synthetic and contains 48,000 primitives. This is a relatively simple model, but the performance of walkthrough is the same without the scene complexity. Figure 7 shows a sequence of images rotating from one point and Figure 8 shows novel views generated by translating along the arbitrary direction. To prevent the dense selected sample, we permit a reasonable error value as 1.85. So we are quite willing to improve the reconstruction quality by adjusting the lower quality threshold. In the circumstances, the initial images are reduced to 21 reference images from the proposed sampling methods.

Fig. 7. Sequence of images rotating from one point.

Fig. 8. Sequence of images by translating along the arbitrary direction.

V. Conclusion

In this paper, we provided a new method of decremental sampling for determining the optimal triangulated data set. Compared with the traditional incremental sampling, the proposed decremental sampling has more practical considerations under the non-uniform sampling circumstances. Especially, considering that the general 3D image warping based reconstruction is achieved by referring the multiple views, we have established a new framework on decremental sampling for this purpose with a triangle mesh optimization mechanism. We were able to reduce the sampling rate by 85% with respect to the capture rate of every couple of inches in experiment without sacrificing the quality of the warped images.

Acknowledgment

This research was financially supported by the Ministry of Education, Science Technology (MEST) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Regional Innovation.

References

[1] D. Aliaga, T. Funkhouser, D. Yanovsky and I. Carlbom, “Sea of images,” IEEE Visualization, pp.331-338, 2002.
[2] S. Fleishman, D. Cohen-Or and D. Lischinski, “Automatic camera placement for image-based modeling,” Computer Graphics Forum, 2000.
[3] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Co., San Francisco, 1979.
[4] This reference is removed for a blind reviewing.
[5] M. Levoy and P. Hanrahan, “Light field rendering,” ACM SIGGRAPH, pp. 31-42, 1996.
[6] W. Mark, L. McMillan and G. Bishop, “Post-Rendering 3D Warping,” Symposium on Interactive 3D Graphics, pp.7-16, 1997.
[7] L. McMillan and G. Bishop, “Plenoptic modeling: An image-based rendering system,” ACM SIGGRAPH, pp.32-43, 1995.
[8] L. McMillan, An image-based approach to three-dimensional computer graphics, Ph.D. thesis, UNC, 1997.
[9] V. Popescu, Forward Rasterization: A reconstruction algorithm for image-based rendering, PhD thesis, University of North Carolina at Chapel Hill, 2001.
[10] H. Schirmacher, W. Heidrich and H.P. Seidel, “Adaptive acquisition of lumi-graphs from synthetic scenes,” EUROGRAPHICS, 1999.
[11] H. Shum and L. He, “Rendering with concentric mosaics,” ACM SIGGRAPH, pp. 299-306, 1999.
[12] C. Zhang, On sampling of image-based rendering data, Ph.D. thesis, Carnegie Mellon University, 2004.