FREE-VIEWPOINT IMAGE SYNTHESIS BASED ON NON-UNIFORMLY RESAMPLED 3D REPRESENTATION

Huei-Yung Lin and Yu-Hua Xiao

Department of Electrical Engineering
National Chung Cheng University
168 University Rd., Min-Hsiung, Chiayi 621, Taiwan

ABSTRACT

Free-viewpoint video is to enable the viewer to choose arbitrary viewpoints for the scene captured by a multi-view imaging system. In this paper we present a multi-layered variate-resolution sampling technique for 3D scene representation. The 3D point cloud obtained from the reconstructed 3D scene is used for novel view rendering. For any given viewpoint, image synthesis with different level of detail is carried out using the quadtree based non-uniformly sampled 3D data points. Experimental results are presented using the 3D model of a reconstructed real object.

Index Terms— 3D representation, novel view synthesis

1. INTRODUCTION

Compared to the rapid growing of the computer and communication technologies, the television and motion picture industry does not make considerable progress in terms of the content production and representation. Most of the existing media formats, ranging from video playback and broadcasting to teleconferencing and interactive video gaming, only provide the 2D images of the real 3D world. The specific viewpoints of the video depends on the camera’s acquisition positions, and cannot be directly controlled by the users. Therefore, future developments of multimedia systems and applications will focus on the creation of interactive and immersive video contents [1], especially for the capability of freely choosing the interested viewpoints.

The objective of this work is to synthesize the images of arbitrary viewpoints for a scene captured by a multi-camera system. This so-called free-viewpoint video generally means that the user is able to assign an arbitrary viewpoint for interactive display. The images captured by a virtual camera with any position and orientation have to be synthesized from the physical cameras based on their locations and the imaging geometry. Although there are several possibilities for the arrangement of multi-view imaging devices, only the convergent camera configuration is suitable for acquiring the surrounding information of an interested scene. In this case, multiple inward-facing cameras are installed around the scene, which can be used for both image acquisition and 3D model reconstruction.

Most of the existing methods for free-viewpoint video generation are based on the novel view synthesis from available images. The commonly used techniques include image-based rendering and model-based scene reconstruction [2]. The former is a geometry free approach, and requires a large number of images for scene interpolation [3]. The latter, on the other hand, only needs a few images for novel view synthesis, but the full geometric scene reconstruction has to be carried out first [4]. To take advantage of the available image compression techniques, it is desirable to have an image-like data format [5]. Furthermore, the imaging system with a dense camera configuration is usually not possible for practical applications. Thus, we will focus on a hybrid geometry and image based rendering method for novel view synthesis—the geometric model of the scene is created first, the image-based rendering is then applied to synthesize the novel view on the fly [6].

In this work, a novel free-viewpoint image synthesis approach is proposed. Based on the 3D model of the scene recovered by shape from silhouettes and multi-view stereo techniques [7], a compact multi-layered data map representation with color and geometric information is constructed through global resampling. To increase the number of 3D points on the regions with sparse sampling in terms of Euclidean measurement, a variate-resolution technique using quadtree structure is carried out on each layer of the data map. For any given viewpoint, image synthesis with different level of detail is then implemented using the non-uniformly sampled 3D data points. Experimental results are presented using a reconstructed real object.

2. 3D MODEL REPRESENTATION

Given the reconstructed 3D model of a scene, a single multi-dimensional array representation can be obtained through a
global resampling process. More specifically, the geometry and texture information of the scene can be represented using the azimuth and zenith angles in a polar coordinate system, with possibly the origin at the centroid of the scene. Thus, a two-dimensional data map with each sample point represented by \( (v, \theta, \phi) \) can be generated, where \( (\theta, \phi) \) is the sampling ray direction and \( v = (\rho, r, g, b) \) is a 4-vector containing the distance and color information.

For a complex 3D scene, the global resampling ray from the origin and along a given direction might intersect more than one point on the 3D surface of the scene. Thus, the representation using polar coordinates could result in multivalued entries. To cope with this problem, a representation with multiple layers of data map is proposed. Each layer of the representation consists of the geometry and texture information. The multi-layered data map (MLDM) is constructed by stacking a number of \( (v, \theta, \phi) \) maps generated according to the sampling order in the \( (\theta, \phi) \) space. That is, for the sampling ray along any given direction, the data point derived from the \( i \)-th intersection with the 3D surface of the scene is recorded on the \( i \)-th layer of the representation map. From the construction of the multi-layered data map, the representation possesses the property that the number and locations of non-empty entries in each data map depends on the topology of the 3D surface.

One major issue on uniform resampling in the \( (\theta, \phi) \) space is that the sampling density in the Euclidean coordinate system is inversely proportional to the distance to the world origin. For a scene scattered around the 3D space, this might result in non-uniform sampling of the 3D surface and lose details far away from the coordinate origin. To overcome this problem, a distance-dependent non-uniform resampling technique is proposed. The basic idea is to increase the sampling density near the directions where the distance between the 3D surface and the coordinate origin is greater than a threshold, while keeping the lower sampling resolution for the surface close to the origin. Since the non-uniform sampling can be carried out on each layer of the data map, the 3D scene can be described by a multi-layered variate-resolution data map (MLVRDM) based on the MLDM representation. For novel view image synthesis, the sampling density can be determined adaptively according to the viewing direction and zooming factor of the virtual camera.

In general, the sampling resolution in both the \( \theta \) and \( \phi \) directions will be doubled locally wherever the data points are not contained in the sphere within the thresholding distance. Thus, there will be \( 4^i \) extra samples generated for the surface point \( v(\theta, \phi) \) located outside the \( i \)-th sphere with radius \( d_i \). In the implementation, the extra sampling directions are processed for each layer of the multi-layered data map using a quadtree structure. Let the initial sampling resolution in the azimuth and zenith directions be \( \Theta \) and \( \Phi \), then the corresponding sampling rates in these two directions are \( 2\pi/\Theta \) and \( \pi/\Phi \), respectively. Suppose a sampling ray with the direction \( (\theta, \phi) \) intersects the 3D surface at the point \( (v, \theta, \phi) \), then the distance between the intersection and the world origin can be written as \( \rho(\theta, \phi) \). The sampling density associated with this direction is increased sequentially by comparing \( \rho(\theta, \phi) \) with \( d_i \)’s, for \( i = 1, 2, 3, \ldots \).

3. NOVEL VIEW IMAGE SYNTHESIS

In the multi-layered data map representation, a colored 3D point cloud is obtained from global resampling in the polar coordinate system. Novel view image can be generated by image interpolation using the projection of the 3D points on the image plane. However, since only the visible surface of the scene can be used for image synthesis, hidden point removal with respect to the virtual view should be carried out first. Furthermore, the 3D model consistent adjacency information among the projected image points should be also considered when synthesizing the surface patch.

Given the 3D point cloud of a scene, to remove the hidden points with respect to a viewpoint, the surface topology has to be established. That is, the 3D mesh model of the scene should be constructed and used for self-occlusion detection. For the multi-layered data map representation, the scene is represented by a 3D volume \( (\theta, \phi, i) \) with the \( (\theta, \phi) \) adjacency information on the same layer. However, the correct surface topology for a complex scene might involve the
transition between layers as illustrated in Fig. 1(a). Thus, the mesh generation has to take the pixel-wise binary discrepancy between the layered data maps into account.

A one-dimensional example of the connection between vertices for a fixed angle $\phi$ is illustrated in Fig. 1(a). It can be considered as a $(\theta, i)$ slice extracted from the 3D volume $(\theta, \phi, i)$. The objective is to find a closed path to represent the original contour by connecting all of the vertices on the $(\theta, i)$ slice. An algorithm for mesh generation on fixed $\phi$ is given in Algorithm 1. Starting from the first layer, the connection between the adjacent vertices in the $\theta$ direction is established if their Euclidean distance is less than a threshold. Otherwise, the path is pointed to the next layer on the same $\theta$ and then reverse the $\theta$ direction for vertex linking. Since it is possible that the correct topology consists of consecutive angle change but on different layer (e.g., link $(8, 3)$ to $(9, 1)$ in Fig. 1(b)), the connecting vertex on the layer discontinuity is determined by the smoothness check. Finally, the mesh generation of a globally resampled 3D scene can be derived from the connection information on the $\phi$ and $\theta$ data slices.

Algorithm 1 Mesh Generation from MLDM

1: $i = 1, \theta = 0, \theta' = \theta + 1, n = 0$
2: while $\mathcal{M} \neq \emptyset$ do
3: \hspace{1em} if $|\rho(\theta', i) - \rho(\theta, i)| < d_1$ then
4: \hspace{2em} connect $(\theta, i)$ and $(\theta', i)$
5: \hspace{2em} $\mathcal{M} \leftarrow \mathcal{M} - v(\theta, i)$
6: \hspace{2em} $\theta \leftarrow \theta + (-1)^n$
7: \hspace{2em} $\theta' \leftarrow \theta' + (-1)^n$
8: \hspace{1em} else if $|\rho(\theta, i + 1) - \rho(\theta, i)| < d_2$ then
9: \hspace{2em} connect $(\theta, i)$ and $(\theta, i + 1)$
10: \hspace{2em} $\mathcal{M} \leftarrow \mathcal{M} - v(\theta, i)$
11: \hspace{2em} $i + +$
12: \hspace{2em} $n + +$
13: \hspace{2em} $\theta' \leftarrow \theta - 1$
14: \hspace{1em} else
15: \hspace{2em} find $j$ such that $|\rho(\theta', j) - \rho(\theta, i)| < d_1$
16: \hspace{2em} connect $(\theta, i)$ and $(\theta', j)$
17: \hspace{2em} $\mathcal{M} \leftarrow \mathcal{M} - v(\theta, i)$
18: \hspace{2em} $\theta \leftarrow \theta + (-1)^n$
19: \hspace{2em} $\theta' \leftarrow \theta' + (-1)^n$
20: \hspace{1em} $i + j$
21: end if
22: $\theta \leftarrow \theta \mod 360$
23: $\theta' \leftarrow \theta' \mod 360$
24: end while

Once the 3D mesh model of a scene is constructed from the point cloud, the hidden points with respect to a certain viewpoint can be detected by forward projection. Let $o$ be the origin of the virtual view, and $(v_i, v_j, v_k)$ forms a triangular mesh of the surface. Then the vertex $v$ should be removed if it is located further than the surface patch, and bounded by

Fig. 2. The point cloud models from global resampling in the polar coordinate system. (a) The reconstructed 3D model of a real object. (b) The first layer data points from uniform resampling. (c) The multi-layered point cloud model. (d) The point cloud model with multi-layered variate-resolution.

the viewing cone determined by the rays $ov_i, ov_j$ and $ov_k$. Since the vertices and connections are ordered from global resampling, all the triangular meshes are verified exclusively to remove the hidden points $v$ based on the following conditions:

(i) $|v| > \min\{||v_i||, ||v_j||, ||v_k||\}$,

(ii) $v_i v_j \times v, v > 0, v_j v_k \times v_j v > 0, \text{and} v_k v_i \times v_k v > 0$.

It should be noted that the triangular meshes with only one or two vertices removed are recorded as partially occluded. They will be preserved for image synthesis, although the hidden points are not shown in the point cloud rendering.

To synthesize the novel view image, the triangular meshes of the 3D scene are projected to the viewpoint with hidden surface removal. Since only the vertices from global resampling contain the color information, they are used to fill out the interior pixels of the projected meshes using bilinear interpolation. For any partially occluded mesh, the corresponding image patch is synthesized first and then overlaid by the occluding one(s). Applying this point-based approach on sufficiently small patches, the synthetic image from any viewpoint can be generated. However, the rendering quality will generally degrade if the projected 2D mesh appears too large.
in the image due to the low \((\theta, \phi)\) spatial sampling or close-up viewing settings.

In the multi-layered variate-resolution data map representation, high density 3D points generated with quadtree structure are used for smooth surface description. Thus, if the projected image points are not close enough for interior region interpolation, the quadtree sampled points are incorporated for novel view image synthesis. The level of quadtree structure required for image interpolation depends on the viewpoint setting and how detailed the initial mesh model is. It varies with the synthesized image region and can be determined by setting a quality threshold based on the size of the projected meshes. By setting the level of quadtree structure adaptively throughout the image plane, a roughly uniform distribution of the projected image points is obtained for constant image-wide rendering quality.

4. EXPERIMENTAL RESULTS

The proposed non-uniformly resampled 3D representation and free-viewpoint synthesis have been carried out on the 3D models of real objects. Fig. 2(a) shows the reconstructed 3D model of a toy object used in the experiment. The color point cloud model obtained from the first layer of uniform resampling is shown in Fig. 2(b). The sampling resolution is \(90 \times 180\) in the \(\theta\) and \(\phi\) directions (i.e. every 2 degrees). It can be seen that the surface topology is incorrect (for example, the front side of the hat) due to the multi-valued mapping in the polar coordinate system. Furthermore, the point density is lower for the surface further away from the coordinate origin located inside the object. Fig. 2(c) shows the point cloud model obtained from the multi-layered uniform resampling. In this case, the complete 3D surface is fully processed, and the local sparseness can be improved using variate-resolution data map as the result shown in Fig. 2(d).

The mesh model generated from the multi-layered data map is shown in Fig. 3(a). It is mainly used to detect the visible data points for arbitrary viewpoint image synthesis. For the image synthesis from a specific viewpoint, the creation of variate-resolution mesh models is generally not required. The quadtree structure can be carried out locally to increase the resolution of the viewable 3D surface. Figs. 3(b) and 3(c) show the novel view synthesis using MLVRDM and MLDM, respectively. For comparison, the same viewpoint rendered from the texture 3D model is shown in Fig. 3(c).

5. CONCLUSION

The capability of arbitrarily choosing the viewpoints is an important feature for many applications. In this work, a novel free-viewpoint image synthesis technique based on multi-layered variate-resolution data map is proposed. Using the non-uniformly sampled 3D data points, image synthesis can be carried out with different level of detail specified by the required rendering quality. Experimental results are presented using the 3D model of a real object. Currently the novel view image is synthesized using the projected data points with the bilinear interpolation, more sophisticated rendering techniques will be investigated in the future.

6. REFERENCES