Improving Quality of Free-Viewpoint Image by Mesh Based 3D Shape Deformation

Satoshi Yaguchi †‡ Hideo Saito †

† Department of Information and Computer Science, Keio University
3-4-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522 JAPAN

‡NTT COMWARE Corporation, NTT Shinagawa TWINS Annex Bldg.
1-9-1 Konan, Minato-ku, Tokyo 108-8019 JAPAN
E-mail yagu@ozawa.ics.keio.ac.jp, saito@ozawa.ics.keio.ac.jp

ABSTRACT

In this paper, we present a method to synthesize high-quality virtual viewpoint image targeting the detailed texture objects. About 30 images are taken from multiple uncalibrated cameras around the object, and the Visual Hull model is reconstructed with Shape from Silhouette method. To deform 3D surface model that is converted from Visual Hull Model using the information such as image texture and object silhouette, the difference between the real object and the reconstructed model is evaluated as a cost function of optimization problem.
Our deforming model algorithm is based on single vertex iterative shifting. The vertex of surface triangle mesh is moved to the selected candidate point that maximizes the cost function. The cost function is consisted by four constraint criteria, texture correlation, smoothness, object silhouette, and mesh shape regularity. In addition to the cost function, such as judging mesh direction and combining / dividing meshes are applied for refined 3D models to avoid mesh folding and mesh size unevenness. The refined model provides a quite accurate dense corresponding relationship between the input images, so that high quality image can be synthesized at virtual viewpoint.
We also demonstrate the proposed method by showing virtual viewpoint images to applying the real image that are taken from multiple uncalibrated cameras.

Keywords

Shape-from-Silhouette, the Visual Hull, shape refinement, Image Based Rendering, weakly calibrated multiple camera system

1. INTRODUCTION

The acquisition of 3D geometric information and generating virtual viewpoint images from multiple cameras are studied many years and still researched actively. The studies of those methods are categorized into two basic methods, correlation-based stereo approach and the Visual Hull [Lau94a] model based approach.

The Multiple Baseline Stereo method [Oku93a, Nay98a] enables acquisition of 3D geometric information with correlation based stereo method from multiple cameras. The advantage of correlation-based approaches is able to handle concave regions. In contrast, it is not stable to handle occluded region. Occluded region causes losing pixels on synthesized virtual viewpoint images. Baselines between each camera are not short enough to remove occluded region, even if there are several tens cameras [Sai03a].

On the other hand, the Visual Hull model based approaches mainly reconstruct a 3D geometric information with shape from silhouette method [Mat00a]. These approaches can synthesize virtual viewpoint image without losing pixels even in partially occluded regions. However, the difference between the real object and reconstructed model may cause blurring in the synthesized virtual images. It is mainly caused by concave region or insufficient number of cameras to curve a Visual Hull model. Especially to apply an object with a detailed texture, the blur greatly effects image quality in appearance. Therefore the quality of the synthesized virtual

Journal of WSCG, ISSN 1213-6972, Vol.14, 2006
Plzen, Czech Republic.

viewpoint image is degraded by inaccuracy of the reconstructed shape.

Recently, the advanced approaches that take into account the both advantages are intensively studied. Those approaches improve the image quality to refine 3D geometric shape with various information such as a texture correlation. The Space Curving Method [Kut00a, Sla02a] removes unnecessary voxels of the voxel-represented model using texture information by reducing difference between constructed models. The techniques for optimizing 3D model by deforming the vertex of surface triangle mesh based on the correlation of the texture have also been proposed [Eck04a, Nob03a].

We have been studied acquisition of 3D geometric information and generating virtual viewpoint images from multiple cameras. We have already proposed a new framework for the Visual Hull based virtual viewpoint image synthesizing method expanding to weakly calibrated multiple cameras using the "Projective Grid Space (PGS)"[Sai99a] in which the coordinates are defined by epipolar geometry between cameras instead of strong calibration [Yag02a, Yag04a]. In the framework, virtual viewpoint images can be synthesized with Image Based Rendering using correspondence map derived from the model like a view morphing method [Sei96a]. However, there is still remained the blurring problem, because the difference between the reconstructed model and the real object is not well considered.

The goal of this paper is to improve a quality of virtual viewpoint image that is synthesized with the framework. The proposed method in this paper aims to reduce the blurring effect to deform the reconstructed model with the information such as image texture or object silhouette. Quite accurate dense correspondence map between the images that is derived from the refined model enables synthesizing high quality virtual viewpoint images without the blur.

2. PROPOSED METHOD

The approach of the proposed method is based on the "Projective Grid Space (PGS)" framework [Sai99a], which relates the 3D object space with 2D image coordinates. Using PGS framework, reconstructing 3D shape model with Shape from Silhouette method and generating virtual viewpoint images becomes possible from uncalibrated multiple cameras [Yag04a].

Input images are taken by uncalibrated cameras that observe around a target object, and silhouette images are synthesized from those images. Several correspondence points between each image are extracted from feature points, and fundamental matrices, which are used for relating 3D object space and 2D image plane with the PGS framework, are calculated from those points. (Please refer the *appendix* or those papers, the details of the PGS framework is shown.)

An initial voxel model is reconstructed by the shape-from-silhouette method on the PGS. By applying the Marching Cubes algorithm, the Visual Hull model is converted to the surface representation model. The vertex of surface triangle mesh is moved to the selected candidate point maximizes the cost function that is consisted by four constraint criteria.

The cost function can be computed only on 2D image domain according to the projection computation from PGS to every image. Obtaining a dense correspondence map between images from the optimized model, high quality virtual view images are synthesized as the image interpolating two input images.

3. SHAPE REFINEMENT

This section describes proposed refinement technique of the 3D-surface model by using only 2D image domain information. Model shape is refined by moving each vertex of surface triangle mesh independently. Each vertex of the surface model is visited sequentially, and then cost function is evaluated at initial position of the vertex. If the evaluated value is under threshold, candidate points of refinement vertex are defined, and each cost value is calculated at those candidate points respectively. The vertex is moved to the candidate point that maximizes the cost function. When all vertexes are visited N times, or cost function is over threshold at all vertexes, shape refinement is finished.

3.1 Vertex Position Optimization

Our algorithm of deforming 3D shape is based on shifting a single vertex iteratively. Each vertex of reconstructed model is visited respectively and shifted independently. The process of shifting vertex position is performed by selecting and moving the candidate vertex to the point that maximizes cost function. The candidate vertexes are defined every iterating cycle.

To simplify the algorithm, the candidate points are defined on the line passing through the target vertex v_0 and a point g that is defined as the center of all adjacent vertexes x_0 , x_1 , ..., x_m . Then $2n+1$ candidate points v_n , **...,** v_{-n} , (within the target vertex) are defined outside and inside of the model surface at intervals of a unit vector scaled by a weight s (see Figure 1).

The parameter s is decided dynamically depending on the value of the cost function of v_0 to enable detailed search, and that is ranged $0.1 \leq s \leq 1.0$. The number of candidate point n is decided depending on the distance between v_0 and g to preserve initial shape detailed structures.

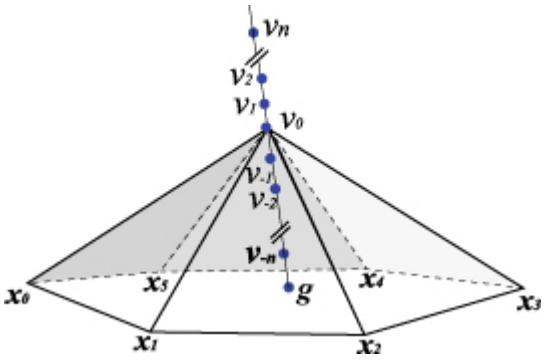


Figure 1. Deform of vertex: Moving candidate points are defined on a line v_0 to *g*.

Cost function V_{-n} , V_{-n+1} , V_0 , …, V_n of each candidate point v_n is calculated by the following equation (1).

$$V_n = \alpha \cdot V_{corr}(v_n) + \beta \cdot V_{smooth}(v_n) + \lambda \cdot V_{shape}(v_n) + \delta \cdot V_{sil}(v_n) \quad (1)$$

- V_{corr} : Texture correlation
- V_{smooth} : Smoothness
- V_{shape} : Triangular shape regularity
- V_{sil} : Silhouette

where α , β , γ , δ , ε are weighting coefficients. Definition of each criterion is described next.

3.1.1 Texture Correlation

Texture correlation of a vertex is determined by the texture of its adjacent triangle meshes. The correlations between each image are calculated by normalized cross correlation to apply all pixels of the mesh. The correlations are defined for the input images from which the vertex and the adjacent meshes are able to observe.

Vertexes of triangle mesh v_{0i} , x_{0i} , x_{1i} in the image *i*, and v_{0j} , x_{0j} , x_{1j} in the image *j* are related to 3D position by fundamental matrices. Inner point *p* of triangle mesh in the image *i* is corresponded to the point in the image *j* by affine transform A_{ij} (see Figure 2).

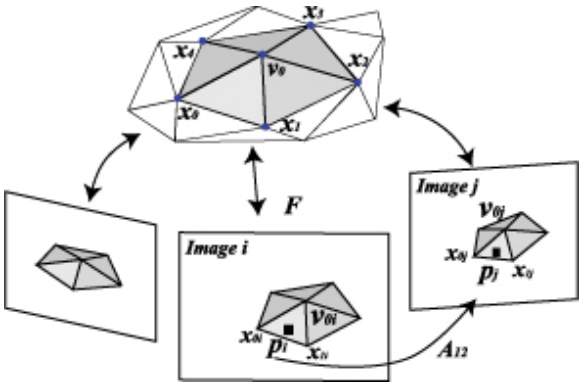


Figure 2. Texture correlation: Adjacent triangle Meshes are projected on images by fundamental matrixes. Each pixel of mesh is corresponded other images by Affine transformation matrixes.

The correlation of v_n of adjacent mesh *k* between image *i* and image *j* is shown equation (2).

$$V_{corr_{kij}}(v_n) = \frac{\sum_p (W_{p_i} - w_i)(W_{p_j} - w_j)}{\sqrt{\sum_p (W_{p_i} - w_i)^2 \sum_p (W_{p_j} - w_j)^2}} \quad (2)$$

where W_{p_i} and W_{p_j} are the color of the point p_i and p_j, respectively, w_i , w_j are the averages of color of all pixel in the image *i* and *j*, respectively.

$V_{corr_{kij}}(v_n)$ is also calculated for all combination of pairs of images on which the vertex v_n is projected onto the pair images without occlusion. Then total $V_{corr}(v_n)$ is computed as the average of all $V_{corr_{kij}}$ calculated for the vertex v_n . $V_{corr_{kij}}(v_n)$ is ranged $-1 < V_{corr}(v_n) < 1$.

3.1.2 Smoothness Constraint

Though surface of the object should be locally smooth, and be continuous, we apply to following smoothness constraint. The constraint is defined depending on the distance $d(v_n)$ between the vertex v_n and *g*, which is determined by all adjacent vertexes of the target vertex as described in **3.1.** In order to reduce over-smoothing, the distance $d(v_0)$ between *g* and the point v_0 , which is the 1/6 distance point on the line segment between v_0 and *g*, is subtracted from $d(v_n)$. Thus, we apply to following function as the smoothness constraint.

$$V_{smooth}(v_n) = \left(d(v_n) - d(v_0)/6\right)^2 \quad (4)$$

where *d* is the distance between v_n and *g*, and weighting coefficient β will be negative.

3.1.3 Silhouette Constraint

The 3D-surface model after deformed should be filled over the initial Visual Hull sufficiently. The vertexes that determine the Visual Hull are constrained to the boundary of the initial silhouettes. However, because the Visual Hull model doesn't accurately express the real contour, for instance, the concave regions, the vertex projected on the boundary of the silhouette is only a candidate determining the contour. The silhouette constraint keeps the refined model to form the initial silhouette.

Therefore the following silhouette constraint is applied to the target vertex and the refinement candidate points (see Figure3).

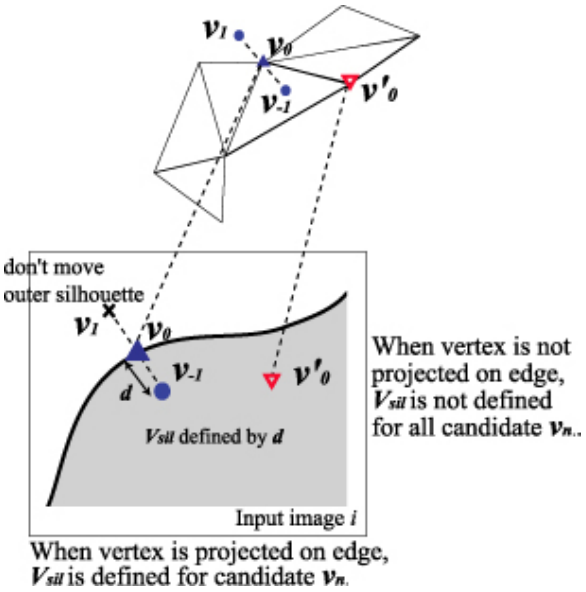


Figure 3. Silhouette Constraint: An image in that vertex projected on edge, silhouette constraint criterion is defined depending on the distance from edge.

1. When the target vertex v_0 is not projected onto the boundary of the silhouette in input image *i*, $V_{sil\,i}$ is not defined ($V_{sil}(v_n)$ of all candidate points are decided to 0.).

2. When v_0 is projected onto the boundary of the silhouette on input image *i*, $V_{sil\,i}$ is defined in proportion to the distance between the projected coordinate of candidate points v_n and that of v_0.

In addition,

3. If v_n is projected on outer silhouette even by one input image, v_n is excluded from candidate.

After all, $V_{sil}(v_n)$ of the candidate point is defined by the sum of $V_{sil\,i}(v_n)$ as following equation (5).

$$V_{sil}(v_n) = \sum_i V_{sil\,i}(v_n) = \sum_i d_i \quad (5)$$

3.1.4 Constraint on Triangular shape regularity

As iterative vertex deforming process goes on, mesh shape and size are changed. If triangle of mesh is very small or very narrow, its texture is not able to derive enough to calculate correlation. So it is preferable that triangle shape is kept shape regularity and size as same as possible.

Therefore, the criterion of constraints on the triangular shape regularity $V_{shape}(v_n)$ of adjacent mesh *k* of vertex v_n is shown by equation (6) [Eck04a].

$$V_{shape\,k}(v_n) = 2\sqrt{3} \cdot \frac{\left|\vec{e}_1 \times \vec{e}_2\right|}{\left|\vec{e}_1\right|^2 + \left|\vec{e}_2\right|^2 + \left|\vec{e}_3\right|^2},$$

$$0 < V_{shape} \le 1 \quad (6)$$

where $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ are edge vectors. This function represents the geometric quality of triangle area rations

The criterion $V_{shape}(v_n)$ of vertex v_n is defined the average of $V_{shape\,k}(v_n)$.

3.2 Additional Constraints

Going on the iterating process, vertex shift may cause mesh folding or mesh size unevenness. Mesh folding changes the topology of the model or the visibility of the vertex, the correlation of the mesh is not able to calculate accurately and the mesh is not rendered correctly (such as “hole” or “overlap” texture is appeared in the synthesized image).

Mesh size variation causes inequality of the precision of cost function, and very small meshes cause invalidate vertex because of having not enough areas deriving textures. In order to avoid such folding or unevenness of the mesh size, following additional constraints are also considered.

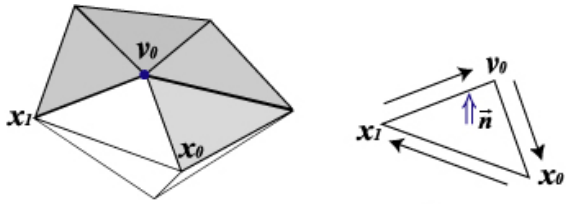
3.2.1 Avoiding Mesh Folding

Mesh folding is occurred when the mesh is not occluded turns to be occluded. It is stated differently, it is caused when the vertex is moved over the boundary of the adjacent mesh.

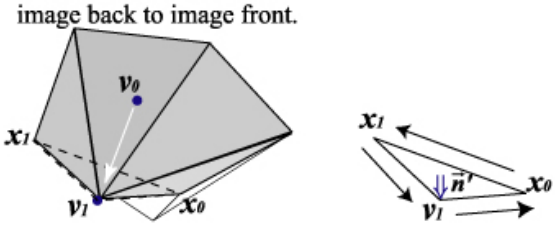
To avoid such mesh folding, normal direction of mesh on the projected image can be used. Vertexes

of mesh v_0 , x_1 , x_2 are indexed in the order of clockwise on the projected image, if the mesh is not occluded, as shown in Figure 4(a). Then the direction of surface normal of the all the meshes adjacent to v_0 are calculated with edge vector. According to the direction, every mesh is determined if it is on the right side or wrong side.

If one of the meshes is not on the right side, mesh is considered as folding as shown in Figure 4(b). Such candidate points of v_0 are excluded from the vertex searching candidate points that are defined **3.1**.



(a)Initial position: Surface normal vector $\vec{n}$ directs through image back to image front.



(b)Mesh folding : The point that changes direction of $\vec{n}$ is excluded out the vertex shift candidate.

Figure 4. Mesh folding: If the candidate point is decided over the boundary of the adjacent mesh, rotation of vertexes and surface normal vector are changed in reverse.

3.2.2 Merging / Dividing Meshes and Vertexes

The distance between the vertexes is changed after the iterating process. Constraint on triangular shape regularity described **3.1.5** is not able to avoid mesh size unevenness. So, when every iteration cycle is finished, meshes and vertexes are divided or merged according to the distance between adjacent vertexes.

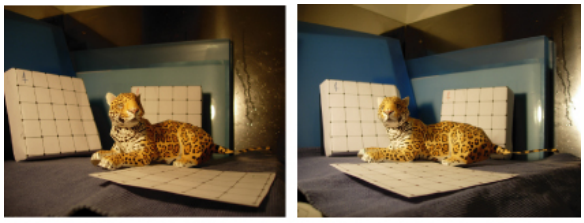
If the distance between adjacent vertexes is over the threshold distance $D_{\max}$, new vertex is inserted to the middle point in these vertexes, and triangles that sharing these vertexes are divide into two triangles each. If the distance between each adjacent vertex is less than the threshold distance $D_{\min}$, these vertexes are merged into one vertex. When two vertexes should be merged, first, the vertex after merged is decided for the middle point of these vertexes. Next, two triangles that share both vertexes are removed, and the mesh index references of each vertex are re-indexed to refer the new vertex.

4. EXPERIMENTAL RESULT

The proposed method has been tested with several real objects. Input images are taken by uncalibrated camera as color images (640×480 pixels BMP format). Results for two real objects, a Jaguar and an elephant are shown in this section.

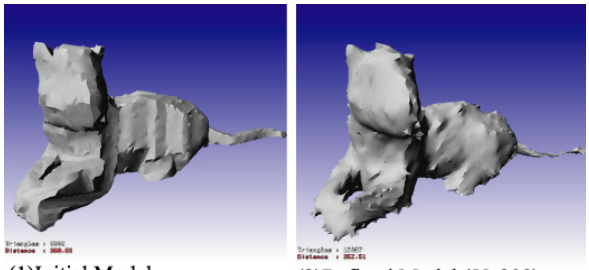
4.1 Jaguar

The target object is a paper craft of "Jaguar" about 20cm×10cm×10cm. 36 images were taken around a target object with a hand-held camera as input images.



(1) input image 1 (2) input image 2

Figure 5. Real Images from that virtual view image shown Figure 7 are synthesized to interpolate.



(1)Initial Model (2)Refined Model (*N*=300)

Figure 6. Initial surface model and refined model iterated *N*=300.

Figure 5 shows the example of the input image. An initial surface model reconstructed in the PGS and the refined model with iteration *N*=300 by proposed method are shown in Figure 6.

Virtual viewpoint image were synthesized with interpolation ratio 5:5 of two real images as shown in Figure 7. Here, Figure (1) and (2) represent the image synthesized from initial model, and the image synthesized from refined model iterated *N*=300, respectively.

In the image of initial model (Figure 7 (1)), the surface texture of the jaguar is blurred. The background area image (blue area) is also rendered as the surface texture of the object. Those bad rendering effects are caused by inaccurate shape of the initial model. Especially in the white ellipse area, the discontinuous pattern is seen, that is because of wrong shape of the jaguar's foot. On the other hand, such bad rendering effect is completely removed in refined model image (Figure 7 (2)).

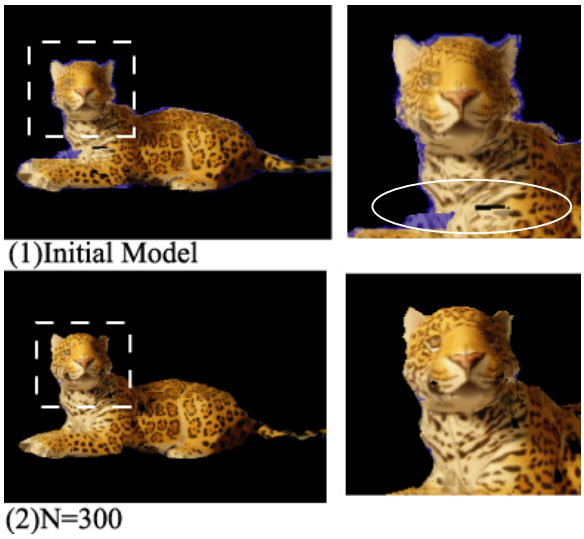


Figure 7. Virtual view image

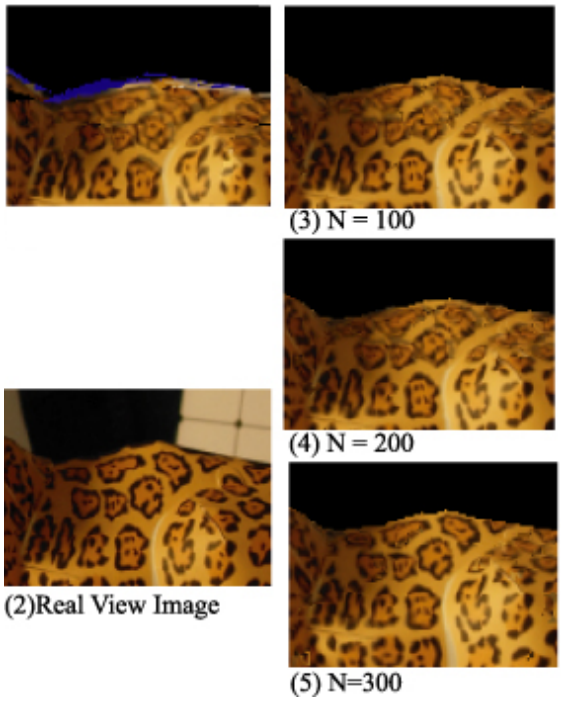


Figure 8. comparison of textures: (1) Rendering from initial 3D shape model. (2) Real image texture about the same viewpoint as virtual viewpoint. (3), (4), and (5) Rendered images with refined model of iteration *N*=100, *N*=200, and *N*=300, respectively.

Next, Figure 8 shows the comparison of texture for each iterating number, initial model, *N*=100, *N*=200, *N*=300, and real viewpoint image. It is understood that the image quality of synthesized image has been improved as the number of iterating process is increased. The real image shown in Figure 8 (2) is taken from about same viewpoint as virtual viewpoint. By comparing *N*=300 image (Figure 8 (5)) with real viewpoint image, about the same quality texture is acquired with refined model.

4.2 Elephant

The target object is an "elephant" about 20cm×20cm×20cm. 30 images were taken as input images. The examples of input images are shown in Figure9.

Figure 10 shows an initial surface model and the refined model with iteration *N*=450 by proposed method. Figure 11 shows the synthesized free viewpoint images. Images on either side are reference real images and middle images are synthesized changing interpolation-weighting factor. Upper images were synthesized from initial model, and lower were synthesized from refined model iterated with *N*=450. In the upper images, many background image areas (blur areas) are distributed, and the border of adjacent rectangular patches are blurred. On the other hand, such bad rendering effects are reduced in the images rendered with the refined model. In this way, free viewpoint images with improved quality texture as same as input images are able to synthesized using proposed method.

Figure 9. Examples of input images.

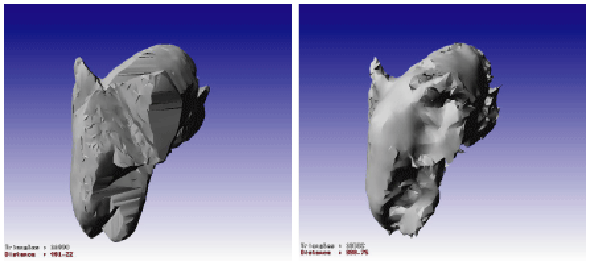


Figure 10. Initial surface model and refined model iterated *N*=450.

5. CONCLUSION

We proposed a method for improving the quality of new view image by deforming 3D surface model from uncalibrated multiple cameras. The cost function of deforming 3D shape model is defined only on 2D image domain according to the projection computation from PGS to every image. Deforming 3D shape model with the texture correlation and other constraints reduces the difference between the reconstructed model and the real object.

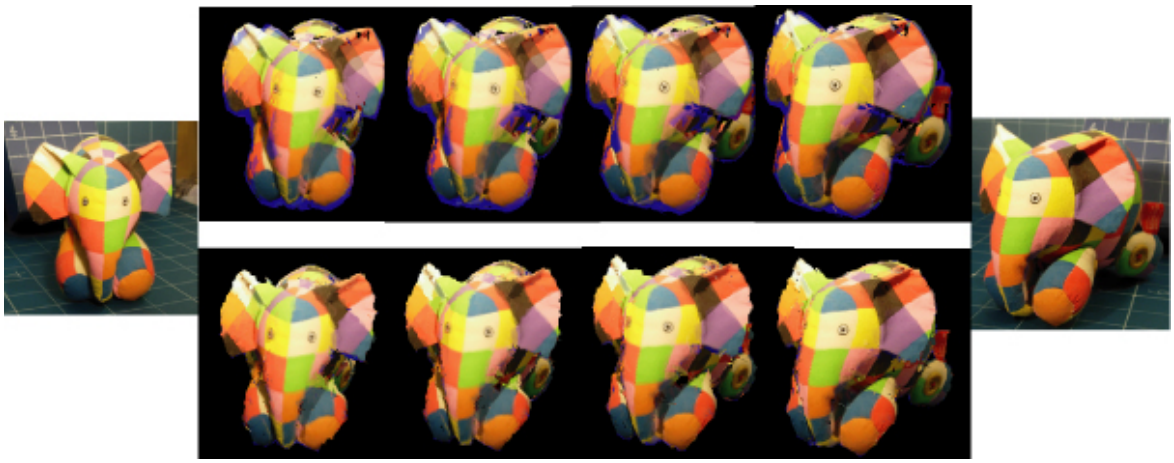


Figure 11. Synthesized free viewpoint images. Images on either side are reference real images, the middle upper images were synthesized from initial model, and lower were synthesized from refined model iterated with *N*=450.

Using refined model removes the blur on the image texture and enables synthesizing high quality free viewpoint image. Our method requires only about 20 to 30 images and does not require strong calibration, so that it is easy to use our method with simple camera systems. For the future work, we will extend this proposed method to the dynamic events and synthesized free viewpoints video from uncalibrated simple camera systems.

6. REFERENCES

[Eck04a] Eckert, G., Wingbermuhle, J. and Niem, W. Mesh Based Shape Refinement for Reconstructing 3D-Objects from Multiple Images: The First European Conference on Visual Media Production (CVMP04), Mar.2004.

[Che93a] Chein, S. and Williams, L. View interpolation for image synthesis: Proc. SIGGRAPH '93, pp.279-288, 1993.

[Cur96a] Curless, B. and Levoy, M. A Volumetric Method for Building Complex Models from Range Images: Proc.of SIGGRAPH '96, 1996.

[Kut00a] Kutulakos, K.N. and Seitz, S.M. A Theory of Shape by Space Carving: IJCV(38), No.3, pp.199-218, 2000.

[Lau94a] Laurentini, A. The Visual Hul Concept for Silhouette-based image understanding: IEEE Trans Pattern Andl. Machine Intell., 16(2), pp150-pp162, Feb. 1994.

[Nay98a] Narayanan, P.J., Rander, P.W. and Kanade, T. Constructing Virtual Worlds using Dense Stereo: Proc. ICCV 98, 1998.

[Mat00a] Matusik, W., Buehler, C., Raskar, R., Gorlter, S. and McMillan, L. Image-Based Visual Hulls: Proc. of SIGGRAPH 2000, 2000.

[Nob03a] Nobuhara, S. and Matsuyama, T. Dynamic 3D Shape from Multi-Viewpoint Images using Deformable Mesh Models: Proc. of 3rd International Symposium on Image and Signal Processing and Analysis, Rome, Italy, September 18-20, pp. 192-197, 2003.

[Oku93a] Okutomi, M. and Kanade, T. A Multiple-Baseline Stereo: IEEE Trans. on PAMI, Vol.15, No.4, pp.353-363, 1993

[Sai99a] Saito, H. and Kanade, T. Shape Reconstruction in Projective Grid Space from Large Number of Images: IEEE Proc. Computer Vision and Pattern Recognition, Vol. 2, pp. 49-54, 1999.

[Sai03a] Saito, H., Baba, S., Kanade, T. Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room: IEEE Trans. on Multimedia, vol.5, no.3, pp. 303-316, Sep. 2003

[Sei96a] Seitz, S. M. and Dyer, C. R. View Morphing: proc. of SIGGRAPH '96, pp. 21-30, 1996.

[Sla02a] Slabaugh, G.G., Schafer, R.W. and Hans, M.C. Multi-Resolution Space Carving Using Level Set Methods: Proc.ICIP02, Vol.II, pp.545-548, 2002.

[Yag02a] Yaguchi, S. and Saito, H. Arbitrary Viewpoint Video Synthesis from Uncalibrated Multiple Cameras: WSCG'2002 - the 10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002, Feb.2002

[Yag04a] Yaguchi, S. and Saito, H. Arbitrary Viewpoint Video Synthesis from Multiple Uncalibrated Cameras: IEEE Trans. on Systems, Man and Cybernetics, PartB, vol. 34, no1, PP.430-439, 2004.

APPENDIX

In our method, 3D point is related to 2D image point without estimating the projection matrices by "Projective Grid Space (PGS)", which can be determined by only fundamental matrices representing the epipoler geometry between two basis cameras. Because the 3D coordinate of PGS is dependently defined by the camera image coordinates, 3D position of any sample points does not have to be measured.

Projective Grid Space

The "Projective Grid Space (PGS)" is defined by camera coordinate of the two basis cameras. Each pixel point (*p*, *q*) in the first basis camera image defines one grid line in the space. On the grid line, grid node points are defined by horizontal position *r* in the second image.

Thus, the coordinate *P* and *Q* of PGS is decided by the horizontal coordinate and the vertical coordinate of the first basis image, and the coordinate *R* of PGS is decided by the horizontal coordinate of the second basis image. Since fundamental matrix F_{21} limits the position in the second basis view on the epipolar line *l*, *r* is sufficient for defining the grid point. In this way, the projective grid space can be defined by two basis view images, of which node points are represented by (*p*, *q*, *r*).

3D-2D Mapping

As described in the previous section, the PGS is defined by two basis views, and the point in the PGS is represented as *A*(*p*, *q*, *r*). The point *A*(*p*,*q*,*r*) is projected onto a_1 (*p*, *q*) and a_2 (*r*, *Y*) in the first basis image and the second basis image, respectively. The point a_1 is projected as the epipolar line *l* on the second basis view expressed as equation (7).

$$l = F_{21}\begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \qquad (7)$$

where F_{21} represents the fundamental matrix between the first and second basis images. The point a_2 is onto *l*, thus the coordinate of a_2 (*r*, *Y*) is decided (see Figure 12).

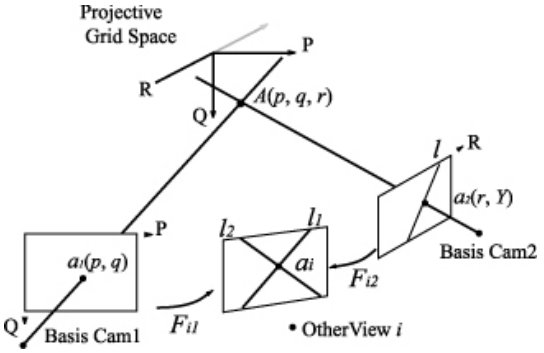


Figure 12. Projection of 3D point onto an image: The point on the projective grid space is projected to the cross point of two epipolar lines in the image of view *i*.

The projected point in *i* th arbitrary real image is determined two fundamental matrices, F_{i1} , F_{i2} between two basis images and *i* th image. The projected point in the *i* th image must be on the epipolar line l_1 of a_1 in the first basis image, which is derived by the F_{i1}. In the same way, the projected point in the *i* th image must be on the epipolar line l_2 of a_2, which is derived by the F_{i2}.The intersection point between the epipolar line l_1 and l_2 is the projected point *A*(*p*, *q*, *r*) in the *i* th image. In this way, every point of the PGS are projected onto every image, where the relationship can be represented by only the fundamental matrices between the image and two basis images.

In this way in the PGS, the fundamental matrices between images determines 3D-2D mapping. In the case of reconstructing 3D shape model with the shape-from-silhouette method, the voxels within a certain region of the PGS are projected into the silhouette image. And the point in an image is able to correspond to other image points to use correspondence map derived from the reconstructed model. And so the correspondence map enables to synthesize new view images as interpolated views and enables to calculate texture correlation between input images.