Relief Mosaic by Joint View Triangulation

Maxime Lhuillier, Long Quan, Harry Shum, Hung-Tat Tsui

Relief Mosaics by Joint View Triangulation

M. Lhuillier[1] L. Quan[2] H. Shum[3] H.T. Tsui[4]

[1] CNRS-INRIA, 655 avenue de l'Europe, 38330 Montbonnot, France
[2] Department of Computer Science, Hong Kong University of Science and Technology
[3] Microsoft Research China, Beijing 100080, P.R. China
[4] Department of Electronic Engineering, Chinese University of Hong Kong

Abstract

Relief mosaics are collections of registered images that extend traditional mosaics by supporting motion parallax. A simple parallax interpolation algorithm based on computed correspondence information allows high quality blur-free and ghost-free mosaics to be created using images from moving hand-held cameras that would not be suitable for traditional mosaicing. The renderer can also display local parallax changes, giving a local but visually convincing illusion of depth. Moreover, relief mosaics can be used for approximate plenoptic modeling from hand-held cameras at lower spatial sampling rates than existing light-field methods. We present a fully automatic correspondence based construction system for relief mosaics, and show how they can be used in applications.

1. Introduction

Image mosaicing has been one of the most effective image-based rendering techniques. The main limitation of previous mosaicing methods [1, 18, 4, 12, 10, 11] is that the registration algorithms assume that the images are free from motion parallax. This implies either that the camera translation is small (fixed viewing position) or that the scene is shallow (near planar). Otherwise, the resulting mosaic has registration errors of order $f \cdot t(d_{min}^{-1} - d_{max}^{-1})$ pixels where f is the focal length in pixels, t the magnitude of the sideways translation, and d_{min} and d_{max} are the near and far scene-camera distances. For deep scenes, special equipment and calibration may be required to fix the viewing position to sufficient accuracy.

In this paper, we propose 'relief mosaics' that extend classical mosaics to allow images with motion parallax. Relief mosaics may viewed simply as collections of registered images with parallax, just as classical ones are collections of images without parallax. The registration of images with parallax is based on view interpolation or view morphing methods [2, 13, 6]. The inclusion of parallax has two important benefits. It allows the creation of high-quality traditional mosaics from parallax containing images taken with freely moving hand-held cameras, and it allows the renderer to produce a local but visually convincing impression of 'relief' or scene depth at very little extra cost.

There have been several previous attempts to deal with the inevitable parallax effects in image mosaicing. Deghosting technique uses local flows of small image patches computed to compensate for the parallax [15]. For dense video sequences, it is also possible to construct 'manifold mosaics' by pasting image stripes from different viewpoints along the camera path [10, 11].

A relief mosaic can be viewed as a '2.5D' plenoptic function [9], intermediate between 2D mosaics, and 3D concentric mosaics [16] and 4D light-fields/lumigraphs [5, 3]. However all of these plenoptic methods use pre-calibrated dense sampling without on-line correspondence, so depth interpolation is still a challenge for finite sampling rates. The static aspect of a relief mosaic is a multi-perspective panorama [19], or a rebinned concentric mosaic [16]. They are also closely related to manifold mosaics [10, 11], however these are constructed from narrow vertical stripes from continuously registered video streams, therefore only limited parallax could be handled and the parallax information is not registered. Relief mosaics are made from fewer images, explicitly register and handle parallax information. We solve the technical challenge of correspondence-based view interpolation for relief mosaic construction using work [6, 7] on the joint view triangulation representation for (real) image interpolation.

More generally, relief mosaics can be used as a building block for more complicated plenoptic modeling to give a kind of generalized light-fields. Such an application is demonstrated in this paper.

The paper is organized as follows. Section 2 describes the relief mosaic representation and its construction from a

linear sequence of real images. Rendering with relief mosaics is presented in Section 3. Applications and results are presented in Section 4 and discussions are given in Section 5.

2. Relief mosaic representation by joint view triangulation

The basic idea of 'relief mosaics' is to assemble images into a composite image using view morphing to cancel their relative motion parallax on the registered overlapping sections, and a heuristic default mapping on the non-overlapping sections to provide visual continuity with the registered ones.

The major technical challenges to using view interpolation or morphing principles [2, 13] for images with parallax are computing dense pixel correspondences and properly handling parallax. Our approach is based on the Joint View Triangulation (JVT) registration and representation method [6, 7]. Globally dense correspondence is seldom achievable in practice, so instead a set of corresponding patches from a 'quasi-dense' correspondence is used as described in [6]. Each patch is a small square in the first image and a quadrangle deformed by a plane homography in the second. A consistent joint view triangulation of the two image planes is constructed based on the corners of the corresponding patches [6, 7]. Consistency for the JVT is defined as follows:

- There is one-to-one correspondence between the matched corner points;
- There is one-to-one constrained Delaunay edge correspondence in the two images. There are two types of Delaunay edge constraints: 1. the external boundary edges of connected and matched patches, and 2. corresponding line segments of polygonal approximations of linked image contour points.

An example of JVT for a pair of video images is illustrated in Figure 1.

A relief mosaic is represented by a collection of images with associated JVT for adjacent pairs of images. Relief mosaics represented this way may be viewed as $2\frac{1}{2}$D plenoptic functions containing a 2D manifold mosaic and a '$\frac{1}{2}D$' local parallax signal. Moreover, relief mosaics can be used as building blocks for more complicated plenoptic modeling, as illustrated in Section 4.

The main advantages of the JVT representation of real images are that it can be computed reliably and that it properly handles occlusions for composition/rendering so long as the occluding objects are sufficiently textured, as it keeps the explicit correspondence information for each triangular patch. It is related to the image impostors used in rendering synthetic large-scale environments [8, 17], however impostors are computed from a single (synthetic) depth image whereas the JVT is a mutually consistent triangulation of two (real) images without depth. It may also be viewed as an image based primitive intermediate between sprites and layered depth images [14].

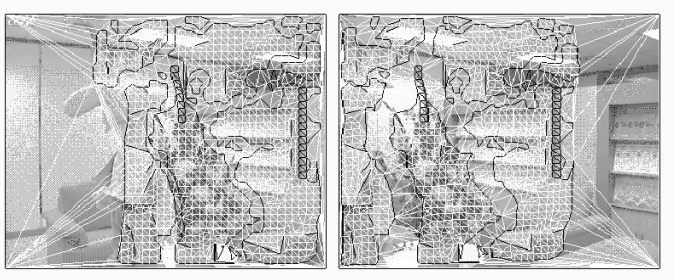

Figure 1. The JVT of the two images. The corresponding patches are inside the black boundaries.

3. Compositing and rendering with relief mosaics

To compose a global static view of a relief mosaic, we warp the non-overlapping sections onto the mosaic coordinates, then locally interpolate the overlapping sections by properly handling the parallax to provide visual continuity with the warped ones.

Two-view compositing Suppose that we are compositing the image pair $I_i(\mathbf{u})$ and $I_{i+1}(\mathbf{u}')$, at a corresponding pair of vertices $\mathbf{u} = (u, v)^T \leftrightarrow \mathbf{u}' = (u', v')^T$ in the JVT.

1. Estimate an approximate global transformation $\mathbf{A}$ using all corresponding vertices in the associated JVT, and use it to map the second image into the first image plane. In our experiments, $\mathbf{A}$ can be either an 8-parameter plane homography or a 2-parameter shift.

   This step is in fact a traditional mosaicing using a global registration method [18].

2. Define an interpolation parameter that varies continuously across the composite image. To maintain visual continuity between different sections, the coefficients for linear interpolation are chosen like this:

$$\lambda(\mathbf{u}, \mathbf{u}') = \begin{cases} 0, & \mathbf{u} \in I_i(\mathbf{u}) \setminus \mathbf{A}(I_{i+1}(\mathbf{u}')), \\ s(\mathbf{u}, \mathbf{u}'), & \mathbf{u}, \mathbf{A}(\mathbf{u}') \in I_i(\mathbf{u}) \cap \mathbf{A}(I_{i+1}(\mathbf{u}')), \\ 1, & \mathbf{A}(\mathbf{u}') \in \mathbf{A}(I_{i+1}(\mathbf{u}')) \setminus I_i(\mathbf{u}), \end{cases}$$

Figure 2. A static relief mosaic at 3700×840 composed from six images at 1004×752.

The function $s(\mathbf{u}, \mathbf{u}')$ should guarantee a smooth transition for the whole composite image, one possible choice might be

$$s(\mathbf{u}, \mathbf{u}') = \frac{d(\mathbf{u}, \mathbf{B}_1)}{d(\mathbf{u}, \mathbf{B}_1) + d(\mathbf{A}(\mathbf{u}'), \mathbf{B}_2)},$$

where $d(\mathbf{u}, \mathbf{B}_i)$ is the distance from $\mathbf{u}$ to $\mathbf{B}_i$, and $\mathbf{B}_i$ is the overlapping section of i-th image border in the composite coordinates.

3. Linearly interpolate each vertex position $\mathbf{u}''$ in the new mosaic coordinates from its two corresponding basis ones $\mathbf{u}$ and $\mathbf{u}'$ by $\mathbf{u}'' = (1-\lambda(\mathbf{u}, \mathbf{u}'))\mathbf{u} + \lambda(\mathbf{u}, \mathbf{u}')\mathbf{A}(\mathbf{u}')$.

4. Warp the first image $I_i(\mathbf{u})$ onto $\tilde{I}_i(\mathbf{u}'')$ in two steps.

   (a) First paint unmatched triangular patches in order of decreasing distortion. A weight w is computed for each warped triangle to measure its inverse distortion degree as the ratio γ of the original triangle area to the warped one, bounded by 1: $w(\mathbf{u}'') = min(1, \gamma(\mathbf{u}''))$.

   (b) Paint the matched triangular patches in heuristic order of increasing disparity over the unmatched ones.

   For a matched triangle $\mathbf{v}_j \leftrightarrow \mathbf{v}'_j, j = 0, 1, 2$, the rendering order is determined as that of the maximum norm of disparity of its vertices: $\max\{||\mathbf{v}_j - \mathbf{v}'_j||, j = 0, 1, 2\}$, assuming that the background moves slower in images than the foreground.

5. Warp the second image $I_{i+1}(\mathbf{u}')$ onto $\tilde{I}_{i+1}(\mathbf{u}'')$ using the same method as for the first image.

6. Create the final synthesized image $I''(\mathbf{u}'')$ by blending the two weighted warped images $\alpha(\mathbf{u}'')\tilde{I}_i(\mathbf{u}'') + (1 - \alpha(\mathbf{u}''))\tilde{I}_{i+1}(\mathbf{u}'')$. The weight $\alpha(\mathbf{u}'')$ is given by $\frac{(1-\lambda(\mathbf{u},\mathbf{u}'))w(\mathbf{u}'')}{(1-\lambda(\mathbf{u},\mathbf{u}'))w(\mathbf{u}'')+\lambda(\mathbf{u},\mathbf{u}')w'(\mathbf{u}'')}$.

Although it is based on heuristic ordering, this two-view compositing algorithm correctly handles disocclusions so long as the occluding objects are textured enough and the depth decreases when the disparity increases. The second part of the algorithm is closely related to [6], but the interpolation parameter is different. It is constant in [6] for a physically valid single perspective view, but variable here for a continuous multi-perspective view. It is also related to McMillan's ordering algorithm [9] using the disparity as a heuristic in place of the true order computed from the camera geometry, which is unknown in our case.

Composing a global static relief mosaic A global static relief mosaic can easily be composed from linear sequences of discrete images taken by a hand-held camera, using the pairwise compositing algorithm described above between adjacent images. We concentrate in this paper on relief mosaics from linear sequences, although more general arrangements are certainly possible. Figure 2 shows an example of a static relief mosaic composed from 6 images.

The static compositing method can be applied directly to create traditional mosaics from images taken with hand-held cameras. Motion parallax not handled by global transformation based alignment methods [18], and lens distortion which needs to be calibrated off-line, are very naturally absorbed by the correspondence based local interpolation of the relief mosaic. The result is a visually high quality mosaic without most of the ghosting (double images) and blurring artifacts due to imperfect global alignment and calibration as showed in Figure 3.

Rendering local parallax changes using a single relief mosaic The static compositing method depends on the choice of the variable interpolation parameter function $\lambda(\mathbf{u}, \mathbf{u}')$. Varying the parameter function produces composite images with different parallax. Locally, the renderer creates an illusion of 'relief', except in the neighborhood of image borders. To keep the global mosaic visually continuous, we may simply replace, among other possibilities, the interpolation parameter $\lambda(\mathbf{u}, \mathbf{u}')$ by $\lambda(\mathbf{u}, \mathbf{u}')^\epsilon$, with $\epsilon > 0$.

For a given viewing window, when $\epsilon < 1$, the parameter λ^ϵ produces a synthetic view closer to the second camera position with ϵ closer to 0. When $\epsilon > 1$ it produces a synthetic view closer to the first camera position with increasing ϵ. Figure 4 shows the local parallax changes rendered from the relief mosaic in Figure 2.

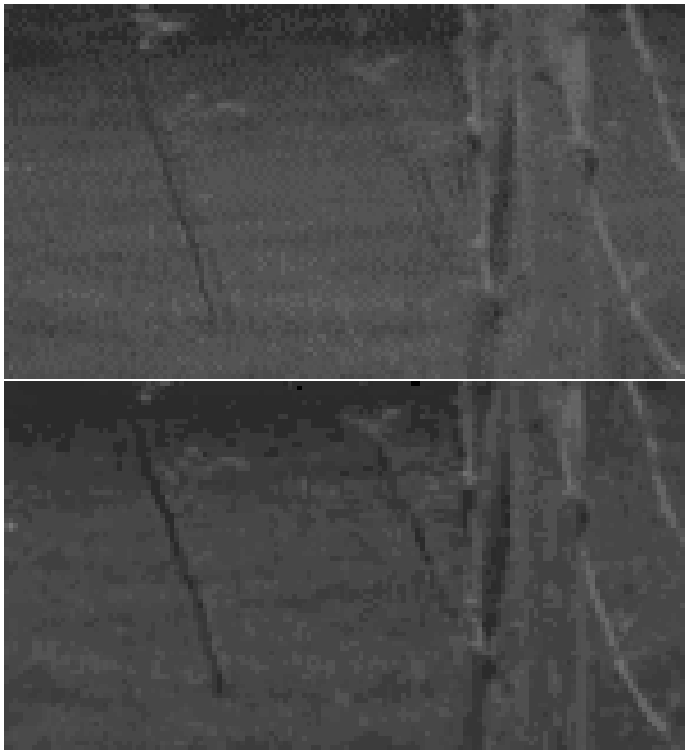

Figure 3. A close-up zoom of a typical traditional mosaic from a hand-held camera. (top) a cylindrical mosaic by a global alignment method; (bottom) a mosaic by relief compositing method. The mosaic on the top is often ghosted due to the parallax of the foreground tree from non-negligible camera motion. The ghosting artifacts are removed using local warping for registration in the relief mosaic.

Figure 4. Local parallax changes rendered at 500×545 resolution from the relief mosaic in Figure 2.

Plenoptic modeling using relief mosaics Only local parallax changes can be rendered with a single relief mosaic. However relief mosaics can be used as building blocks for more general plenoptic modeling of larger environments. This gives a more powerful representation than plenoptic modeling [9] and a more space/construction efficient one than Lightfield/Lumigraph [5, 3] and concentric mosaics [16]. We have implemented a sparse horizontal Lightfield system for outward looking scenes.

The viewing field consists of a roughly regular $M \times N$ grid of images. It can be viewed as N relief mosaics each made from M images. Conceptually we may regard each static relief mosaic as a regular image as illustrated in Figure 5. The correspondence and JVT operations are then performed between composite images. However the correspondence information can not be computed directly from the composite images as they are not single perspective images. Instead, correspondence information is computed separately for each vertical pair (i, j) and $(i, j+1)$ and finally merged in the whole composite images.

Suppose that the i-th and $(i+1)$-th relief mosaics have dimensions $m_i \times n_i$ and $m_{i+1} \times n_{i+1}$, and the viewing window $m \times n$. Let $\mathbf{m} \leftrightarrow \mathbf{m}'$ be corresponding vertices in the two static relief mosaics and (x, y) be the barycentric coordinates of the viewer's position between the two. The position of the vertices in the synthesized view are determined by linear interpolation

$$\mathbf{m}''(y) = (1-y)\mathbf{m} + y\mathbf{m}' - \mathbf{c}(x, y),$$

where

$$\mathbf{c}(x, y) = \begin{pmatrix} x((1-y)m_i + ym_{i+1}) - m/2 \\ (1-y)n_i + yn_{i+1}) - n/2 \end{pmatrix}.$$

This renders a view that moves forward/backward (y) or left/right (x) within the field. The local parallax of each component relief mosaic can also be added to the synthesized views to produce an illusion of local left/right turn. We pre-compute the vertices of each component relief mosaic corresponding to the two extreme positions (right by ϵ_{min} and left by ϵ_{max}), using

$$\mathbf{r} = (1 - \lambda^{\epsilon_{min}})\mathbf{u} + \lambda^{\epsilon_{min}}\mathbf{A}(\mathbf{u}')$$

and

$$\mathbf{l} = (1 - \lambda^{\epsilon_{max}})\mathbf{u} + \lambda^{\epsilon_{max}}\mathbf{A}(\mathbf{u}').$$

Then the position of vertex in the synthesized view is located at

$$\mathbf{m}(\mu) = (1-\mu)\mathbf{l} + \mu\mathbf{r}.$$

The final synthesized coordinates $\mathbf{m}''(x, y, \mu)$ are given by

$$(1-y)\mathbf{m}(\mu) + y\mathbf{m}'(\mu) - \mathbf{c}(x, y).$$

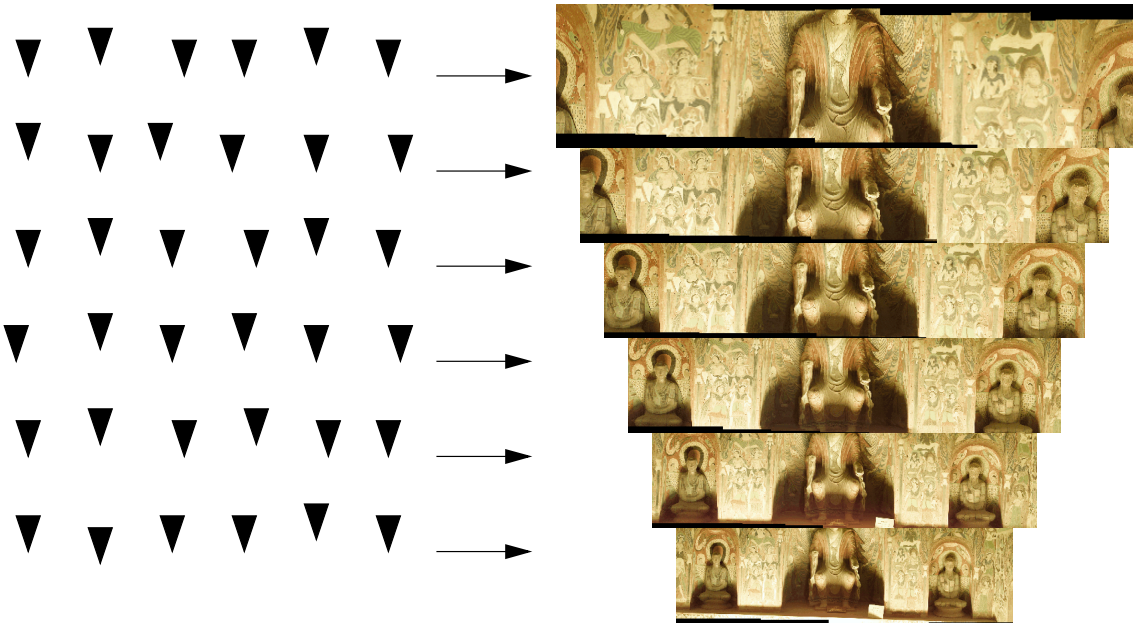

Figure 5. Each row of images are first composited into a relief mosaic for a horizontal lightfield.

4. Results

The office scene was taken by a single camera moving off-line along a circle. The panorama can not be composed using traditional mosaic registration techniques due to its strong motion parallax. The relief mosaic at 3577×288 resolution shown in Figure 7 was created from 36 video images at 360×288 resolution. The significant parallax between adjacent images shown in Figure 1 (the maximum disparity can reach $1/10$ of the image size) makes composition impossible with traditional mosaicing methods. Local parallax changes rendered with this relief mosaic are shown in Figure 6.

Figure 6. Local parallax changes rendered at 210×220 resolution from the relief mosaic in Figure 7.

A grid of 6×6 images at 1004×752 resolution of the wall containing the fresco and statues of Dunhuang cave #285 was captured with a hand-held camera. The grid contains 6 translating fronto-parallel views of the wall taken at 6 different depths. A relief mosaic was generated for each depth as shown in Figure 5. Figure 2 shows one of the static relief mosaics and Figure 4 shows local parallax changes of the central statue by varying ϵ. Traditional mosaicing methods are impossible for these images as the statue has significant depth variation w.r.t. the moving camera. This grid of 6×6 images is used to create a horizontal lightfield described above. Some synthesized views are shown in Figure 8. A synthesized MPEG sequence with a circular viewer path in the horizontal field of the cameras is given in the conference CDROM (the image size is reduced due to storage limitations). The results are globally good, although there are occasional artifacts at abrupt disparity changes.

5. Conclusion

In this paper, we have presented relief mosaics which extend traditional mosaics by supporting motion parallax. They can be viewed as $2.5D$ plenoptic functions. They are composed of collections of images with parallax, for example captured using a moving hand-held camera. Motion parallax is allowed and no camera geometry information is necessary. Relief mosaics can be used to create high-quality traditional mosaics, and also to render local parallax changes. Moreover, relief mosaics can be combined to form more complicated plenoptic-like models, giving a lightfield

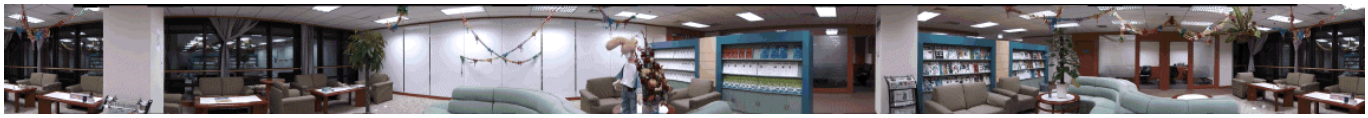

Figure 7. A static relief mosaic at 3577×288 from 36 video images. Note in Figure 1 that significant parallax exists in two adjacent images.

Figure 8. Synthesized images at 700×500 resolution from the field of 36 images at 1004×752 resolution. We may notice that a triangular patch is mis-placed in the front arm in the third view.

captured with a hand-held camera. This has been demonstrated with our system on an example with only 6×6 images. We are currently extending it to produce a more general (real) image based modeling and rendering system.

Acknoledgements

We greatly acknowledge Duanhuang Museum and Duanhuang Foundation for the help of acquiring images used in this paper. We would also like to thank Bill Triggs for many fruitful discussions.

References

[1] S.E. Chen. Quicktime VR - an image-based approach to virtual environment navigation. SIGGRAPH'95.

[2] S.E. Chen and L. Williams. View interpolation for image synthesis. SIGGRAPH'93. 1993.

[3] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. SIGGRAPH'96.

[4] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. *ICCV*'95.

[5] M. Levoy and P. Hanrahan. Light field rendering. SIGGRAPH'96.

[6] M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. *CVPR*'99.

[7] M. Lhuillier and L. Quan. Edge-constrained joint view triangulation for image interpolation. *CVPR*'2000.

[8] P. Maciel and P. Shirley. Visual navigation of large environments using textured clusters. SIGGRAPH'95.

[9] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. SIGGRAPH'95.

[10] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. *CVPR*'97.

[11] B. Rousso, S. Peleg, I. Finci, and A. Rav-Acha. Universal mosaicing using pipe projection. *ICCV*'98.

[12] H.S. Sawhney, S. Ayer, and M. Gorkani. Model-based 2D & 3D dominant motion estimation for mosaicing and video representation. *ICCV*'95.

[13] S.M. Seitz and C.R. Dyer. View morphing. SIGGRAPH'96.

[14] J. Shade, S. Gortler, L.W. He, and R. Szeliski. Layered depth images. SIGGRAPH'98.

[15] H.-Y. Shum and R. Szeliski. Construction and refinement of panoramic mosaics with global and local alignment. *ICCV*'98.

[16] H.Y. Shum and L.W. He. Rendering with concentric mosaics. SIGGRAPH'99.

[17] F. Sillion, G. Drettakis, and B. Bodelet. Efficient impostor manipulation for real-time visualization of urban scnenery. *Eurographics*'97.

[18] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. SIGGRAPH'97.

[19] D.N. Wood, A. Finkelstein, J.F. Hughes, C.E. Thayer, and D.H. Salesin. Multiperspective panoramas for cel animation. SIGGRAPH'97.