

Video Synthesis made simple with the X-Slits Projection

Doron Feldman Assaf Zomet Shmuel Peleg Daphna Weinshall
School of Computer Science and Engineering
The Hebrew University of Jerusalem, 91904 Jerusalem, ISRAEL
email: {doronf,zomet,peleg,daphna}@cs.huji.ac.il

Abstract

We propose novel algorithms for image based rendering (IBR) by simulating a new family of scene-to-image projections, the Crossed-Slits (X-Slits) projections. New X-Slits images can be easily generated from a sequence of images captured by a translating pinhole camera, and those images look compelling and realistic. We show how this can be used for rendering movies of virtual walkthroughs by a very fast and simple mosaicing algorithm, providing a strong 3-D sensation of parallax and occlusions, as well as specular and transparency.

1 Introduction

The X-Slits projection, shown in Fig. 1, involves two slits I_1, I_2 which are two different lines in \mathcal{R}^3 , together with an image plane Π that does not contain any of the slits. For every 3-D point there is a single ray which connects the point with both slits simultaneously. The intersection of this ray with the image plane defines the projected image of the 3-D point. Fig. 1a shows a special case of the X-Slits projection, where the two slits are perpendicular to each other and are parallel to the image plane. We call this special arrangement the Parallel-Orthogonal X-Slits projection (POX).

In this paper we show that virtual X-Slits images can be generated by mosaicing perspective images, or equivalently by slicing the space-time volume¹ of images. Applications include the generation of virtual walkthroughs from a single sequence of images. The advantage of producing virtual X-Slits images is that they are very easy to produce, yet compelling and resemble real perspective. The results show that in many practical cases the discrepancies between X-Slits images and perspective images are hardly noticeable. Another application is object visualization. Specifically, such visualization can be useful when

¹The space-time volume is simply a stacking of the input images, a.k.a. the epipolar volume.

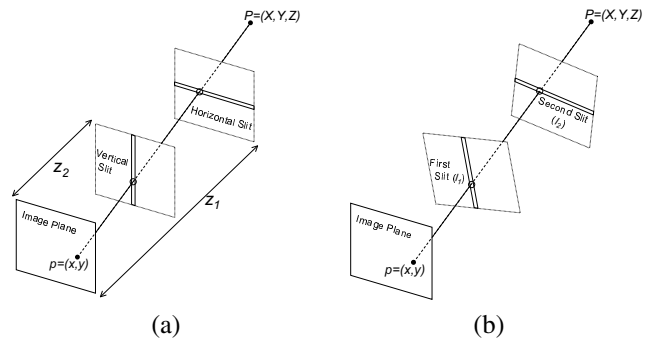


Figure 1. In the X-Slits projection a ray passes through a point and two linear slits. (a) A projection where the slits are orthogonal to each other and parallel to the image plane (POX-Slits camera). (b) A general projection, with two arbitrary slits I_1, I_2 .

one wants to see simultaneously different sides of an object. Both applications are done without recovering the scene or object geometry.

There has been much work on non-perspective projection models [13] and the generation of non-perspective images from video sequences [14, 15, 19, 8]. While these images were mostly used as a visual summary of the video, or for 3D visualization, the present work makes the generation of virtual walkthroughs possible.

Image-based rendering (IBR) was also treated extensively, where rays from a set of input images are collected and a new image is rendered by resampling the stored rays [12, 5, 11]. There were attempts to make IBR more efficient and more general [4, 1].

The present work is mostly related to [16, 17] with several differences: First, rather than trying to approximate the perspective projection, we accurately define the projection geometry of the resulting images, and analyze the model limitations. Second, the rendering tool is very simple - slicing of the space-time volume. Consequently the most important feature of our technique is the fact that ray-sampling for the generation of new views does not require detailed accounting of the parameters of the generating im-

ages. As we show below, if the camera's motion is sideways and constant, every vertical planar slice of the space-time volume gives *some* valid X-Slits image.

Curiously, a X-Slits device was designed in the 19th century by one of the inventors of color photography, Louis Ducos du Hauron [10], under the name "transformisme en photographie".

2 The X-Slits Projection Geometry

Consider the camera configuration as shown in Fig. 1.b. The *projection ray* of a point $\mathbf{p} \in \mathcal{P}^3$ should intersect the two camera slits $\mathbf{I}_1, \mathbf{I}_2$. The image of \mathbf{p} is the intersection of this projection ray with the image plane. Furthermore, all points on a projection ray will project to the same image point. It can be shown that the projection of a 3D point \mathbf{p} by a X-slits camera is given by:

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} \propto \begin{pmatrix} \mathbf{p}^T \mathbf{S}_1 \mathbf{Q}_1 \mathbf{S}_2 \mathbf{p} \\ \mathbf{p}^T \mathbf{S}_1 \mathbf{Q}_2 \mathbf{S}_2 \mathbf{p} \\ \mathbf{p}^T \mathbf{S}_1 \mathbf{Q}_3 \mathbf{S}_2 \mathbf{p} \end{pmatrix} \quad (1)$$

and \mathbf{p} is projected to $(\frac{x}{w}, \frac{y}{w})$. $\mathbf{S}_1, \mathbf{S}_2$ are the dual-Plücker matrices representing the two slits: For every point $\mathbf{p} \in \mathcal{P}^3$, $\mathbf{S}_i \mathbf{p}$ is the plane passing through \mathbf{p} and the i -th slit. $\mathbf{Q}_1, \mathbf{Q}_2$ are the Plücker matrices representing the image plane axes, and \mathbf{Q}_3 is the Plücker matrix representing the line at infinity that lies on the plane. A full derivation and of the X-slits projection appears in [2].

2.1 Rotating the Image Plane

What is the relation between two X-Slits images which are generated by the same slits, but different projection planes Π and Π' ? Note that the set of rays intersecting the image plane are determined by the two slits. Thus, there exists a mapping between the two image planes which is invariant of the 3D structure of the viewed scene, similarly to the case of rotations in pinhole cameras.

Specifically, we derive the relation between the corresponding points on Π' and Π , which are the projection of the same scene point \mathbf{p} (see example in Fig. 1). Let $\mathbf{m} \in \mathcal{P}^3$ denote a point on plane Π , and let $\mathbf{j}, \mathbf{k} \in \mathcal{P}^3$ denote two points at infinity which also lie on plane Π . Clearly the projection of \mathbf{p} on Π can be written as $x\mathbf{j} + y\mathbf{k} + w\mathbf{m}$, with image homogeneous coordinates (x, y, w) . Similarly, let (x', y', w') on Π' correspond to $\mathbf{q}' \propto x'\mathbf{j}' + y'\mathbf{k}' + w'\mathbf{m}'$. By definition \mathbf{q} lies on the planes $\mathbf{S}_1 \mathbf{p}$ and $\mathbf{S}_2 \mathbf{p}$, and therefore \mathbf{q} is also projected to (x', y', w') . Denoting

$$\mathbf{M} = (\mathbf{j} \quad \mathbf{k} \quad \mathbf{m}), \quad p = \begin{pmatrix} x \\ y \\ w \end{pmatrix}$$

it follows that $\mathbf{q} = \mathbf{M}p$. Projecting $\mathbf{M}p$ on Π' :

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} \propto \begin{pmatrix} p^T \mathbf{M}^T \mathbf{S}_1 \mathbf{Q}'_1 \mathbf{S}_2 \mathbf{M} p \\ p^T \mathbf{M}^T \mathbf{S}_1 \mathbf{Q}'_2 \mathbf{S}_2 \mathbf{M} p \\ p^T \mathbf{M}^T \mathbf{S}_1 \mathbf{Q}'_3 \mathbf{S}_2 \mathbf{M} p \end{pmatrix} \quad (2)$$

Note that \mathbf{M} depends only on plane Π , \mathbf{S}_1 and \mathbf{S}_2 depend only on slits \mathbf{I}_1 and \mathbf{I}_2 , and $\mathbf{Q}'_1, \mathbf{Q}'_2, \mathbf{Q}'_3$ depend only on plane Π' .

The relation between the two image points $p = (x, y, w)^T$ and $p' = (x', y', w')^T$ is therefore not a homography as in the pinhole case, but a quadratic projective transformation.

2.2 A Special Case: POX-Slits projection

In the POX-Slits projection (Fig. 1a) the slits are perpendicular and parallel to the image plane. The 3D to 2D projection in this case is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -Z_1 \frac{X}{Z-Z_1} \\ -Z_2 \frac{Y}{Z-Z_2} \end{pmatrix} \quad (3)$$

These projection equations are similar to the bi-centric model analyzed in [18]. In this projection a straight line in the scene is mapped to a hyperbole in the image.

A special case of POX-Slits cameras is the linear pushbroom camera, in which the vertical slit resides on the plane at infinity. This camera was investigated in [6], where it was shown that a line in 3D is projected to a hyperbole.

The most visible distortion in POX-Slits images is the variation of aspect-ratio (especially in pushbroom images). From (3) it follows that an object at depth Z with aspect-ratio of 1 would appear on the image plane to have an aspect-ratio of $\frac{\Delta y}{\Delta x} = \frac{Z_2}{Z_1} \cdot \frac{Z-Z_1}{Z-Z_2}$. In practice, we found this distortion to be rather insignificant. If we cancel the aspect-ratio distortion for some intermediate depth value Z_0 by scaling, the aspect-ratio distortion for other objects would typically not exceed 10%.

A change in viewing direction can be simulated by rotating the image plane. The image transformation can be computed by applying the POX-slits configuration to Eq. (2). Specifically, we fix the vertical slit at $Z = 0$, the horizontal slit at $Z = Z_2 - Z_1$ and the parallel plane at $Z = -Z_1$ (the POX-Slits configuration, translated to align the vertical slit with the Y axis). Defining the rotated plane by three axes $\mathbf{m}' = (m_1, m_2, m_3, 1)^T$, $\mathbf{j}' = (j_1, j_2, j_3, 0)^T$, $\mathbf{k}' = (k_1, k_2, k_3, 0)^T$, we get the following transformation from (2):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -Z_1 \frac{j_1 x' + k_1 y' + m_1}{j_3 x' + k_3 y' + m_3} \\ -Z_2 \frac{j_2 x' + k_2 y' + m_2}{j_3 x' + k_3 y' + m_3 + Z_1 - Z_2} \end{pmatrix} \quad (4)$$

This 2-D transformation allows us to readily produce the images corresponding to oriented image planes Π , given

the corresponding POX-Slits image with parallel image plane Π' .

3 Mosaicing New Views

Virtual X-slit views can be generated from an input sequence captured by a pinhole camera translating along a horizontal line. In the new X-Slits image, the horizontal slit is fixed, corresponding to the path of the camera. On the other hand the vertical slit can be positioned in a rather large area, where the exact location of the vertical slit is determined by the mosaicing process. A virtual walkthrough can be obtained by moving the vertical slit to different locations, and synthesizing for each location a new X-Slits image.

A virtual X-slit image is generated as follows:

- From each image t , sample a column (or a vertical strip) according to the line sampling function $s(t)$, whose parameters are determined by the desired location of the vertical slit. $s(t)$ is described in Section 3.1.
- Paste the columns together into a single image, as in “regular” mosaicing.
- Warp the final image to account for X-Slits image-plane orientation, using Eq.(4).

3.1 Column Sampling for Mosaicing

Let our input be a sequence of images captured by a pinhole camera translating in constant speed along the X axis from left to right. A new panoramic image can be generated by pasting columns from the input images, see schematic illustration in Fig. 2, a top-down view. In Fig. 2 a sequence of positions of the real pinhole camera is shown, together with the corresponding field of view. The moving input camera, whose optical center is located at positions $\mathbf{c}(t) = (X_t, 0, 0)$, generates images according to the following mapping:

$$(X, Y, Z) \implies \left(f \frac{X - X_t}{Z}, f \frac{Y}{Z} \right) \quad (5)$$

We denote the range of columns (x) in each pinhole image as $[-r, r]$, and the range of camera pinhole positions (X_t) as $[-l, l]$ (see Fig. 2). The range of columns in the synthesized image is $[-(r+l), r+l]$. For each $t \in [-1, 1]$, we assign to the $(l+r)t$ column of the new image the image values at the rt column of the pinhole camera positioned at $(lt, 0, 0)$ (i.e., $X_t = lt$). It now follows from Eq. (5) that $rt = f \frac{X-lt}{Z}$. In addition, for each column $x \in [-(r+l), (r+l)]$ in the new image, $t = \frac{x}{l+r}$ and therefore

$$x = \frac{l+r}{r} f \cdot \frac{X}{Z + f \frac{l}{r}}$$

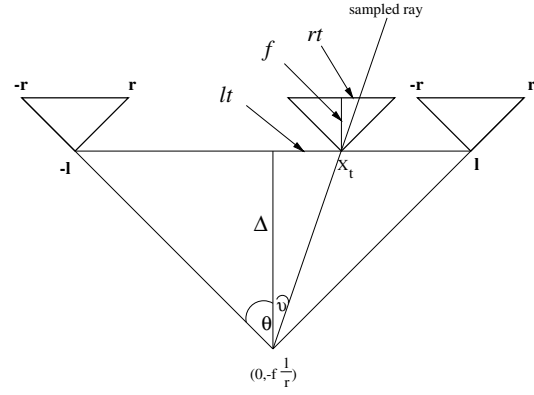


Figure 2. Top view of column sampling.

It follows that the image obtained by such sampling corresponds to the projection given by

$$(X, Y, Z) \implies \left(f_x \frac{X}{Z + \Delta}, f_y \frac{Y}{Z} \right) \quad (6)$$

where $f_x = \frac{l+r}{r} f$ is the horizontal focal length, $f_y = f$ is the vertical focal length, and $\Delta = f \frac{l}{r}$ is the distance between the two slits. Observe that this defines a POX-Slits projection with a vertical slit at $Z = -f \frac{l}{r}$ (see Fig. 2) and a horizontal slit at $Z = 0$ (which is the path of the translating pinhole camera).

Suppose next that instead of taking the rt column from the camera at location $(lt, 0, 0)$, we choose an arbitrary linear column sampling function. More specifically, for sampling function

$$t = \alpha s + \beta, \quad (7)$$

we take the rs column of the lt camera. (Recall that r, l are fixed, while t, s are free parameters which determine the rate of column sampling). Let the field of view of the original pinhole camera be 2θ . It can be shown that such a choice of columns defines the mapping

$$(x, y) = \left(f_x \frac{X - X_0}{\Delta + Z}, f_y \frac{Y}{Z} \right) \quad (8)$$

where $X_0 = \beta l$, $\Delta = \frac{\alpha l}{\tan \theta}$, $f_y = f$ and $f_x = f + \Delta$.

The generated images are not perspective, but follow the X-Slits projection. To see this, we observe that all rays producing the image must intersect the following two lines: (i) The camera path, since each projection ray must be collected by some camera whose optical center is on this line. (ii) The vertical line located at (X_0, Z_0) (as in Eq. (8), where $Z_0 = Z + \Delta$); this is the vertical slit.

The projection model is therefore defined by a family of rays intersecting a pair of lines (“slits”), projecting 3D points onto a plane. Moreover, the model is POX-Slits (compare Eq. (8) with Eq. (3)).

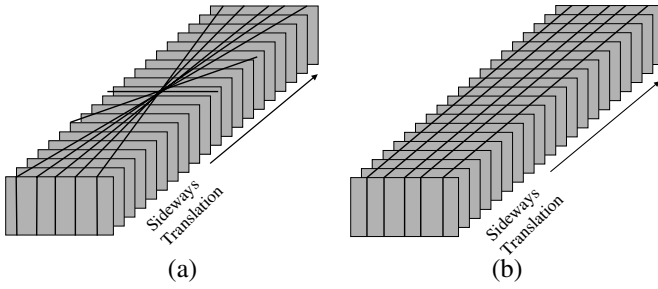


Figure 3. A schematic description of images generated as slices in the space-time volume. (a) Changing the angle of the slice moves the vertical slit inside and outside the scene. (b) The central slice gives a pushbroom image (the “traditional” mosaic). Sliding parallel slices in the space-time volume results in different viewing directions of skewed pushbroom images.

In the derivation leading to (8) we effectively showed that any linear sampling function yields a valid new POX-Slits image. Furthermore, we can set the location of the vertical slit to any (X_0, Z_0) by fixing $\alpha = -\frac{Z_0}{f} \tan \theta$ and $\beta = \frac{X_0}{f}$ in Eq. (7). It can further be shown that even when the internal parameters of the camera are unknown (but fixed), and the camera motion is parallel to the image plane, any linear column sampling results in a X-Slits image. Even when the camera motion is not parallel to the image plane new X-Slits views can be generated, see [2].

3.2 Implementation Issues

Variable Camera Speed: When the camera moves in a linear trajectory but varying orientation and speed, we compensate for this variability by estimating camera motion (see [7]) and by derotating the image planes. Specifically, we compute the 2D rotation and translation between consecutive input frames using the method described in [3], and warp the images to cancel any 2D rotation and vertical translation. The residual 2D translation is used to estimate the rate of 3D camera translation. This approach is similar to the pushbroom mosaicing technique described in [14].

The Space-Time Volume: A useful representation for the visualization of this process is the Space-Time Volume (or the *epipolar volume*), which is constructed by stacking all input images into a single volume. In case of constant sideways camera motion, any planar slice in the volume according to (7) is a X-Slits image. This process is illustrated in Fig. 3.

Rendering new X-Slits images becomes as simple as slicing the space-time volume with a plane (using appropriate interpolation). There is an inverse duality between rotation and translation. Rotating the slice in the space-time volume provides X-Slits images that represent forward-backward translation in the scene. Translating the slice in

the space-time volume results in the rotation of the viewing angle.

4 Extended X-Slits Models

The X-Slits projection presented so far was characterized by two properties: The set of projection rays was defined by two linear slits, and the image surface was assumed to be a plane. This model can be generalized by modifying the form of the slits, or the shape of the image surface, or both. In this section we introduce a few useful generalizations of the basic X-Slits model.

Circular Camera Motion: One of the slits in the synthetic X-Slits images is the camera path. When the camera path is not linear, The synthesized X-Slits view has a curved slit. An interesting family of such views has one circular slit and one linear slit.

One way to generate such images is to use a camera rotating off-axis on a circle, as in concentric mosaics [16]. Concentric mosaics allow the generation of images in which the viewer can move continuously inside a circular region. Each image generated from concentric mosaics is a X-Slits picture: One slit is the horizontal circular path of the camera center, and the second slit is a vertical linear slit placed at the location of the viewer. To generate an image from a different viewing position, the vertical slit is placed in the new location.

While images generated from the concentric mosaics point outward, and the viewer location is inside the circle, it is also interesting to generate inward looking images from locations outside the circle. This can be realized by moving the camera in a circle around an object, or by having a stationary camera viewing an object rotating on a turntable. Now the location of the viewer in the synthesized images can be as far or as close to the object as we wish, inside or outside the circular slit (camera path).

Panoramic Input Images: Synthesizing new panoramic views from an input sequence of perspective panoramic images is also possible. In this case a cylinder can be used as the image surface for both the input and output images, and the novel panoramic views can be generated in a method very similar to the usual X-Slits images.

Specifically, we assume that the cylindrical input images are captured by a panoramic camera moving on a straight line. One slit of the synthesized panoramic view is the path of the input camera. The other slit can have an arbitrary location.

The virtual image can be generated by pasting and scaling columns from the input images. Let one slit be the path



Figure 4. Original images captured by a sideways moving camera. The synthesized movie tracked the person in the scene, with simulated camera rotation and forward motion. Three frames of the synthesized movie are shown.

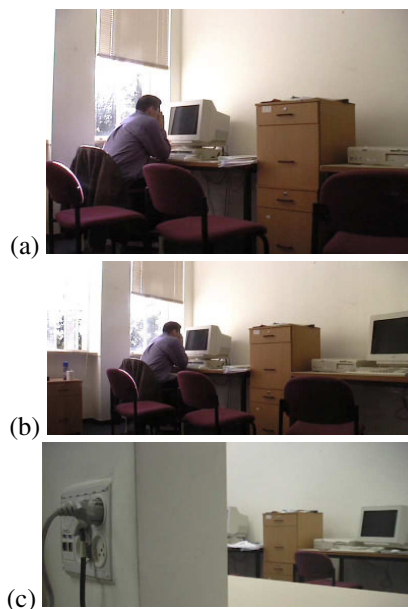


Figure 5. Images from impossible camera positions. A sideways moving camera was used inside a room, and one frame is shown in (a). (b) is a synthesized image in which the vertical slit is located far behind the original track. (c) is a normal (pinhole) picture from the same location as (b), where most of the scene is obscured by a wall and a table.

of the camera, and the second slit parallel to the Y axis and passing through $(X_0, 0, Z_0)$. If we index the columns by horizontal angle, it can be shown that column α should be the same column α , when sampled from the image captured at time $t = Z_0 \tan \alpha + X_0$. Thus new view synthesis can be done simply by stacking the input images in a

space-time volume, and slicing the volume by the function $t = Z_0 \tan \alpha + X_0$. After slicing, every column α should be scaled by $\frac{R}{\frac{Z_0}{\cos \alpha} + R}$, where R is the radius of the cylinder.

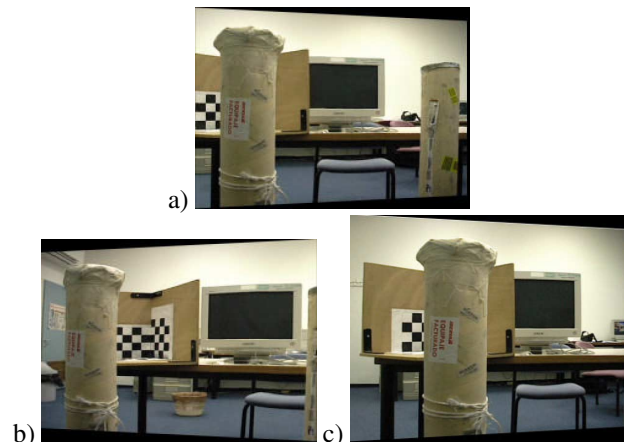


Figure 6. Original sequence was captured by a sideways moving camera. We generated a X-Slits movie where the virtual camera rotates about an object in the scene (a-b), and then translates forward diagonally (c).

5 Experimental Results

Virtual Walkthrough: In Figs. 4-7 we synthesized from a sideways moving camera new sequences which included simulated forward and backward camera motion, as well as rotation, with visible parallax and lighting effects. It is interesting to see that images can be simulated from locations where no regular picture can be taken (Fig. 5b-c). Full synthesized sequences for figures 4-5-6 can be seen in the supplementary file. In Fig. 8 we show new views generated from a sequence of a rotating object, where the new image includes forward motion with parallax.

Object Visualization: Here we demonstrate the use of the X-Slits projection for object visualization. Specifically, we show how an object can be “flattened”, revealing several of its sides simultaneously. This is done by positioning the vertical slit behind the object (Fig. 9). Since the image is a valid X-Slits image that can be characterized and analyzed, we need not worry about such issues as duplicate images, which usually require hand-crafted stitching.

6 Concluding Remarks

We showed that new view generation with the X-Slits projection model is greatly simplified, reducing in many interesting cases to mosaicing or the slicing of the space-time volume. Here the X-Slits theory helps the user to

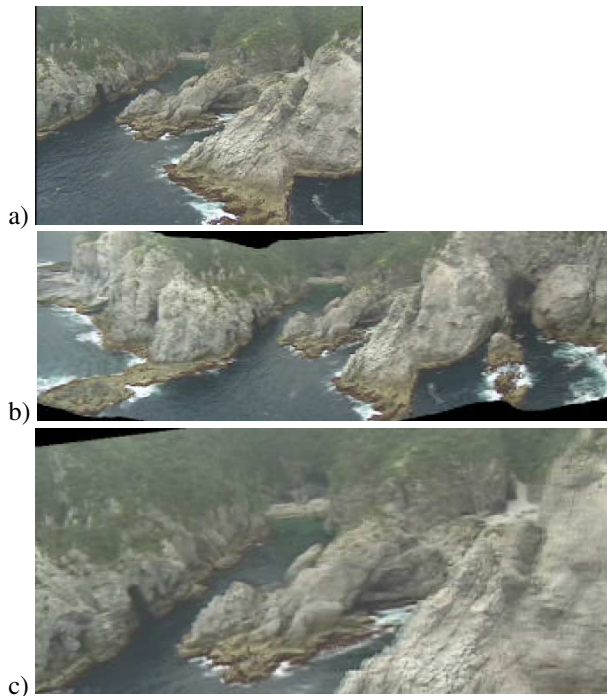


Figure 7. Virtual walkthrough generated from a sequence taken by a freely flying helicopter. (a) A frame from the input sequence. (b)-(c) Two images rendered in forward motion (diagonal slices).



Figure 8. Two synthesized images from a forward moving viewpoint, using as input a sequence obtained by rotating the object. Note parallax around the fence edges.

“drive” the slicing process in order to get the desired effect. This allows us to generate images from “impossible” positions, like behind the back wall of a room or in front of a glass barrier. We also generated movies with new ego motion, such as forward-moving movies from a side-moving input sequence. Although not perspective, the movies generated in this way appear compelling and realistic.

References

[1] D. Aliaga and I. Carlbom. Plenoptic stitching: A scalable method for reconstructing interactive walkthroughs. In *Proc. of ACM SIGGRAPH'01*, pages 443–450, 2001.

[2] Assaf Zomet, Doron Feldman, Shmuel Peleg and Daphna Weinshall. Non-Perspective Imaging and Rendering with the Crossed-Slits Projection. Hebrew University TR:2002-41, July

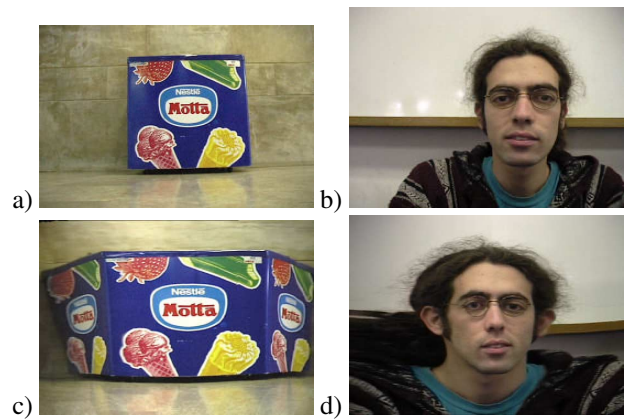


Figure 9. Object representation. a, b) The original input images. c, d) Visualization with the vertical slit located behind the object. The object is seen as if “opened” inside-out, giving a cubist effect: multiple sides are seen in a single picture.

2002. http://leibniz.cs.huji.ac.il/tr/acc/2002/HUJI-CSE-LTR-2002-41_paper.ps.gz

[3] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. of ECCV:237–252*, 1992.

[4] C. Buehler, M. Bosse, S. Gortler, M. Cohen, and L. McMillan. Unstructured lumigraph rendering. In *Proc. of ACM SIGGRAPH'01*, pages 425–432, 2001.

[5] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Proc. of ACM SIGGRAPH'96*, pages 43–54, 1996.

[6] R. Gupta and R. I. Hartley. *Linear Pushbroom Cameras IEEE PAMI*, 9(19):963-975, 1997.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[8] H. Ishiguro and M. Yamamoto and S. Tsuji. Omni-Directional Stereo *IEEE PAMI*, 14(2):257–262, 1992.

[9] S. Kang. A survey of image-based rendering techniques. In *Video-metric VI*, 3641:2–16, Jan 1999.

[10] R. Kingslake. *Optics in Photography*. SPIE Optical Engineering Press, 1992.

[11] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of ACM SIGGRAPH'96*, pages 31–42, 1996.

[12] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Proc. of ACM SIGGRAPH'95:39–46*, 1995.

[13] T. Pajdla. Stereo with Oblique Cameras. *International Journal of Computer Vision*, 47(1/2/3):161-170, 2002.

[14] S. Peleg, B. Rousso, A. Rav-Acha and A. Zomet. Mosaicing on Adaptive Manifolds *IEEE PAMI*, 10(22):1144–1154, 2000.

[15] S.M. Seitz. The Space of All Stereo Images In *Proc. of ICCV'01*, pages I: 26-33, Vancouver, Canada, July 2001.

[16] H. Shum and L. He. Rendering with concentric mosaics. In *Proc. of ACM SIGGRAPH'99*, pages 299–306, 1999.

[17] T. Takahashi, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Arbitrary view position and direction rendering for large-scale scenes. In *Proc. of IEEE CVPR*, pages 296–303, SC 2000.

[18] D. Weinshall, M. Lee, T. Brodsky, M. Trajkovic, and D. Feldman. New view generation with a bi-centric camera. In *Proc. of 7th ECCV*, 614–628, Copenhagen DK, 2002.

[19] J.Y. Zheng and S. Tsuji. Panoramic Representation for Route Recognition by a Mobile Robot *IJCV*, 9(1): 55-76, 1992.