Mosaicing Impossible Stereo Views

Shmuel Peleg[a] Daphna Weinshall[a] Doron Feldman[a] Yael Pritch[b]

[a]School of Computer Science and Engineering
The Hebrew University of Jerusalem, 91904 Jerusalem, ISRAEL
[b]HumanEyes Technologies Ltd., Jerusalem, ISRAEL

ABSTRACT

Most image rendering methods try to mimic real cameras by generating images having the perspective projection. In contrast, a unique power of image mosaicing is the ability to generate new views with "impossible" projections which are not perspective. This can be done with mosaicing methods that construct a panoramic mosaic image by stitching together narrow strips, each strip taken from a different source image. A different selection of strips gives a different mosaicing effect using the same set of source images, including the generation of stereo images.

For example, given a sequence of source images from a camera moving sideways, a set of panoramic stereo views can be generated, even though perspective cameras allow only a very narrow view for stereo images. And even though the original (single) camera moved sideways, a sequence of forward moving stereo images can be generated.

As each of the stereo views is generated synthetically from the original images, stereo effects can be adjusted in the post production stage. Such effects include changing the stereo baseline and the vergence. Post production enables the same set of original images to be used for generating stereo images for various displays and viewing conditions.

1. INTRODUCTION

Perspective projection forms the foundation of stereo imaging, since our eyes, as well as most of our cameras, observe the world through a pinhole. What can we do differently when not limited by the perspective projection, but are free to use other projection models?

In this paper we present the use of an alternative projection model for stereo imaging. Specifically, we propose a projection defined by two slits - the Crossed-Slits (X-Slits) projection. In the X-Slits model, the projection ray of every 3D point is defined by the line that passes through the point and intersects both slits. The image of a point will be the intersection of the projection ray with the image surface. The X-Slits projection is especially effective for image based rendering (IBR); we show examples how stereo X-Slits images can be generated by non-stationary mosaicing from a sequence of images captured by a translating pinhole camera.

We do not propose to build X-slits cameras, but to generate such images by a non-stationary mosaicing of a sequence of regular input images captured by a translating camera. What makes the mosaicing process "non-stationary" is that different columns are sampled from different images, according to a sampling function which depends on the frame index (usually corresponds to time). We focus mainly on linear sampling functions that are easy to implement efficiently, while providing a strong 3-D sensation.

1.1. The X-Slits Camera: Outline

Fig. 1 shows a schematic design of an X-Slits "camera".[1] Specifically, the X-Slits camera has two slits $\mathbf{l}_1, \mathbf{l}_2$ which should be two different lines in $\mathcal{R}^3$, together with an image plane that does not contain any of the slits. For every 3-D point not on either of the slits 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. The camera in Fig. 1 has two slits that are orthogonal to each other and are parallel to the image plane.

E-Mail of authors: {peleg, daphna, doronf}@cs.huji.ac.il, yaelp@humaneyes.com. This work was supported in part by a grant from the Israel Science Foundation, and by the EU under the presence initiative through the BENOGO project.

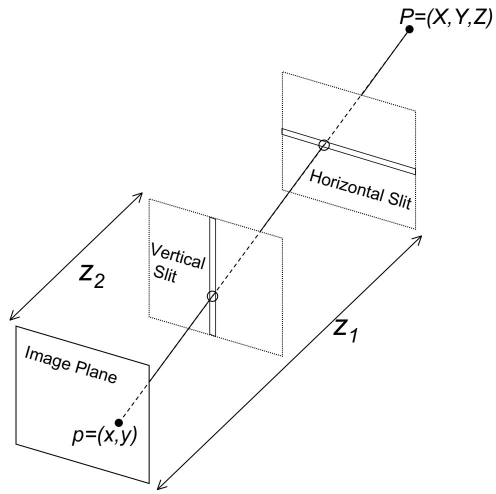

Figure 1. A schematic design of a X-Slits camera where the slits are orthogonal to each other and parallel to the image plane. Z_1 denotes the horizontal focal length and Z_2 denotes the vertical focal length. The projection ray of a 3-D point $\mathbf{p} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is shown, with circles showing its intersection points with the 2 slits.

The X-Slits model is a valid 3-D to 2-D projection, defining a many-to-one mapping from the 3-D world to the 2-D image plane. In Section 3 we discuss how to generate new X-Slits images from images taken by a regular pinhole camera translating along a straight line. Specifically, we show that virtual X-Slits images can be generated by non-stationary mosaicing of perspective images, or equivalently by slicing the space-time volume* of images.

The generation of an arbitrary virtual walkthrough from a single sequence of images is discussed in Section 5. We show a number of examples where we move the virtual stereo camera around and change its orientation; the examples demonstrate realistic changes in parallax and reflections. Moreover, the results show that in many practical cases the discrepancies between X-Slits images and perspective images are hardly noticeable. We note that all this is done by plain mosaicing, without recovering the scene geometry.

This paper includes work done on the theory of X-slits projections[1] and on generation of panoramic stereo images from a camera rotating off-axis.[2]

2. THE X-SLITS PROJECTION

Let us study the configuration in which the slits are orthogonal and parallel to the image plane as in Fig. 1.

We assume that the vertical slit is at $X = 0$, $Z = Z_1$ and the horizontal slit is at $Y = 0$, $Z = Z_2$. It follows that a 3D point (X, Y, Z) is projected into the image point (x, y) as follows:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -Z_1 \frac{X}{Z-Z_1} \\ -Z_2 \frac{Y}{Z-Z_2} \end{pmatrix} \tag{1}$$

X-slits images are different from perspective images in a few ways. In the perspective projection straight lines in the scene are projected to straight lines in the image. This is not the case under the X-Slits projection, where straight lines are projected to conic sections. There is also an aspect-ratio distortion in X-slits images, most notable in pushbroom images.[3] The apparent aspect-ratio of objects in the X-slits image depends on their depth. This is unlike the perspective model, in which the distortion in aspect-ratio is constant for all objects at all depths.

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

To demonstrate the magnitude of the aspect ratio distortion, consider the following example: Suppose the depth range of objects in the scene is $3-5$ meters (measured from the horizontal slit at Z_1, i.e., $3 < Z - Z_1 < 5$), and assume that the images are normalized so that objects at the depth of $3.84m$ appears undistorted (i.e., $Z_0 - Z_1 = 3.84m$). If the vertical slit is behind the horizontal slit at $\Delta = -2.5m$, the aspect-ratio distortion would not exceed 10% from a perspective image.

Most projection models which have been discussed in the literature, with the exception of the oblique camera,[4] are special cases of the X-Slit camera:

- Perspective Projection – this is the case when the two slits intersect; the intersection point is the optical center of the perspective projection.
- Parallel Projection – the two slits are at infinity.
- Linear Pushbroom[3] – the vertical slit resides on the plane at infinity.

3. NON-STATIONARY MOSAICING GENERATING A SINGLE MOSAIC IMAGE

Our goal is to synthesize new X-Slits views from "regular" perspective images. The input sequence is assumed to be captured by a pinhole camera translating along a line in 3-D space in a roughly constant speed, and without changing its orientation or internal calibration. We will show how to generate a new X-Slits image by mosaicing, where the directions of the two slits of the virtual X-Slits camera are as follows:

1. One slit lies on the path of the optical center of the moving pinhole camera.
2. The second slit is parallel to image's vertical axis, and its location is determined by the mosaicing process.

The non-stationary mosaicing is defined as follows:

- From each image t, sample the vertical column whose horizontal coordinate in the image is $s(t)$.
- Paste the columns together into a single image, as in "regular" mosaicing.

Equivalently, we can define that each column s in the mosaic image will be taken from frame t in the input sequence according to the function $t(s)$. The use of $t(s)$ is more convenient than the use of $s(t)$ except in the case when the same column is taken from all images as shown in Fig. 4.b. In this case $t(s)$ is undefined, and only $s(t)$ can be used.

The parameters of the line sampling function $s(t)$ (or equivalently the function $t(s)$) determine the exact location of the second slit of the virtual camera. Using this observation, a virtual walkthrough is the result of generating a sequence of X-Slits images via non-stationary mosaicing, while moving the second slit along a planar path in the plane defined by the path of the virtual camera and the sampled image line.

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. We generate a new panoramic image by pasting columns from the input images, as illustrated in Fig. 2. We start by sampling the left column of the first (leftmost) image, and conclude by sampling the right column of the last (rightmost) image. In between, intermediate columns are sampled from successive images using a linear sampling function.

A schematic illustration of this setup is given in Fig. 3a, in a top-down view. A sequence of positions of the real pinhole camera is shown, together with the corresponding field of view. The moving input camera, whose optical centers are located at positions $\mathbf{c}(t) = (X_t, 0, 0)$, generates images according to the following mapping:

$$\mathbf{p} = (X, Y, Z) \implies p = (x, y) = (f\frac{X - X_t}{Z}, f\frac{Y}{Z}) \tag{2}$$

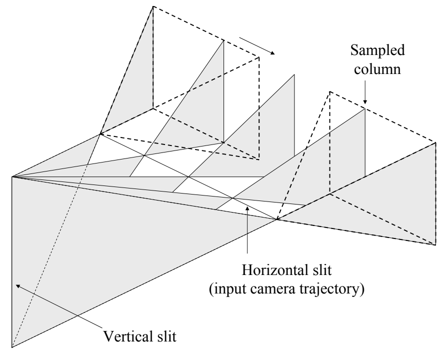

Figure 2. The non-stationary column sampling routine which is used to synthesize new images.

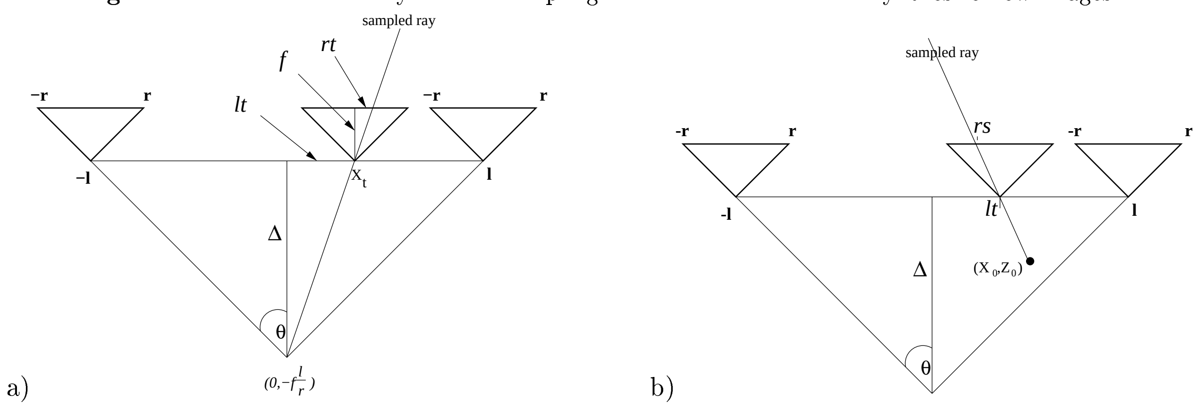

Figure 3. New image formation with two possible positions of the vertical slit (see text).

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. 3a). The new synthesized image is constructed by pasting columns from the input images. 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., $X_t = lt$, see Fig. 3a). It now follows from Eq. (2) 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

$$X = \frac{rt}{f}Z + lt = x\left(\frac{r}{l+r} \cdot \frac{Z}{f} + \frac{l}{l+r}\right)$$

or

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

Observe that this defines a vertical slit at $Z = -f\frac{l}{r}$ (see Fig. 3a). The horizontal slit is at $Z = 0$, same as for each pinhole camera. Eq. (2) therefore becomes the following projection

$$\mathbf{p} = (X, Y, Z) \implies p = (x, y) = (f_x \frac{X}{Z + \Delta}, f_y \frac{Y}{Z}) \tag{3}$$

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.

Suppose next that instead of taking the rt column from the camera at $(lt, 0, 0)$, we choose an arbitrary linear column sampling function. More specifically, for $t(s) = \alpha s + \beta$, we take the rs column of the lt camera, see Fig. 3b. (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

$$(x, y) = ((f + \frac{\alpha l}{\tan\theta})\frac{X - \beta l}{\frac{\alpha l}{\tan\theta} + Z}, f\frac{Y}{Z}) \tag{4}$$

This can be written simply as

$$(x, y) = (f_x \frac{X - X_0}{\Delta + Z}, f_y \frac{Y}{Z}) \tag{5}$$

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

The method described so far produces images which do not follow the perspective projection model, since the focal lengths are not the same vertically and horizontally. They do, however, follow the X-Slits projection model defined above. To see this, we observe that all rays producing the image must intersect the following two lines:

1. The line of camera motion; this is because each projection ray must be collected by some camera whose optical center is on this line.
2. The vertical line located at (X_0, Z_0) (as in Eq. (5), where $Z_0 = Z + \Delta$); this is shown above to be 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.

In the derivation leading to (5) we effectively showed that any linear sampling function yields a valid new X-Slits image. Furthermore, we can set the location of the vertical slit to (X_0, Z_0) by fixing $\alpha = -\frac{Z_0}{l}\tan\theta$ and $\beta = \frac{X_0}{l}$. This result enables us to synthesize new views of the scene with any vertical slit of our choice, by sampling the columns of the original input sequence according to $t(s) = \alpha s + \beta$, with α and β assigned the appropriate values.

It can be shown that when the camera motion is parallel to the image plane, any linear sampling of the columns results in a X-Slits image. This is true even when the internal parameters of the camera are unknown (but fixed).

3.1. Implementation Issues

In this section we discuss what to do when the motion of the camera deviates from constant speed (Section 3.1.1), and how the aspect ratio of the resulting mosaic is determined (Section 3.1.2). We also present an alternative implementation of mosaicing, namely the slicing of space-time volume (Section 3.1.3).

3.1.1. Variable Camera Speed

When the camera moves in a linear trajectory but varying orientation and speed, we can compensate for this variability by estimating camera motion[5] and by derotating the image planes. We found that when the changes in camera orientation are small, a simple approximation is sufficient. Specifically, we compute the 2D rotation and translation between consecutive input frames,[6] and warp the images to cancel any 2D rotation and vertical translation. The residual 2D translation is used to estimate the 3D velocity of the translating camera. This approach is similar to the pushbroom mosaicing technique.[7]

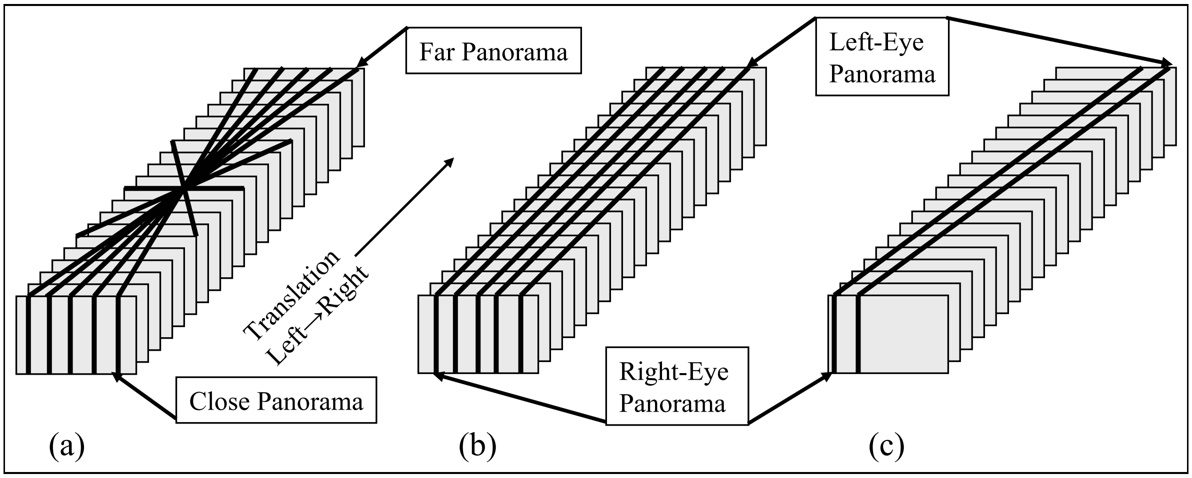

Figure 4. A schematic description of images generated as slices in the space-time volume. a) Changing the orientation of the slice moves the vertical slit forward and backward. b) Pushbroom mosaics. The central slice gives an image with orthogonal rays (the "traditional" mosaic). Sliding parallel slices in the space-time volume results in different viewing directions, or a skewed pushbroom. Any two parallel slices are a stereo pair. c) Any two parallel slices provide a pair of stereo images. images. Rotating the two slices provided a motion forward and backward of the stereo pair.

3.1.2. Aspect-Ratio Normalization

The most apparent aspect of the distortion in X-Slits images is the variation of aspect-ratio. To reduce this distortion, we scale vertically the new images. This normalization to correct the aspect ratio is essential to achieve good results.

Specifically, the distortion on the *image plane* of objects at depth Z can be written as $\frac{Z}{Z+\Delta} \cdot \frac{f_y}{f_x}$ in the notation of Eq. (5). In order to keep the horizontal field-of-view angle constant in the walk-through animation, we sample all the columns from left to right (from the appropriate frames, according to the column sampling function). Without any scaling, this process generates an image in which only the plane at infinity ($Z = \infty$) appears undistorted. Therefore, in order to cancel the distortion at depth Z_0, we must scale the image vertically by the factor:

$$1 + \frac{\Delta}{Z_0} \tag{6}$$

3.1.3. The Space-Time Volume

In Section 3 we described how to synthesize a X-Slits image by sampling columns from the input images. Specifically, column s in the mosaic image is selected from frame t according to the following linear sampling:

$$t(s) = \alpha s + \beta \tag{7}$$

where t denotes the camera translation. Recall that α, β are free parameters which control the location of the vertical slit.

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 a constant sideways camera motion, any planar $t - y$ slice in the volume according to (7) is a X-Slits image. This process is illustrated in Fig. 4; it assumes that the input sequence has a high frame-rate so that simple interpolation of the volume is sufficient. Thus rendering new X-Slits images is as simple as slicing a $t - y$ plane in the space-time volume.

It is interesting to note the inverse duality between rotation and translation in the slicing of the space-time volume. When the input images are from a camera that moves sideways, rotating the slice in the space-time volume as shown in Fig. 4.a provides X-Slits images that represent forward-backward translation in the scene. However, translating the parallel slice in the space-time volume as shown in Fig. 4.b results in the rotation of the viewing direction of a pushbroom camera. This change in the viewing direction gives us the stereo capability.

4. X-SLITS STEREO

Section 3 described the generation of X-Slits panoramas from "regular" perspective images. The input sequence is captured by a camera translating along a horizontal line in a roughly constant speed, without changing its orientation or internal calibration. Panorama synthesis is performed as follows:

- Each column s in the mosaic image is selected from frame t in the input image sequence.

The geometry of the resulting panoramic image is X-Slits, corresponding to a X-Slits camera with one slit parallel to the image vertical axis, and a second slit overlapping the path of the camera. The parameters of the strip sampling function $t(s)$ determine the location of the vertical slit of the virtual camera.

In the simplest case, when the camera is moving sideways in constant speed, we can make the following general observation:

Any linear column sampling function $t(s) = \alpha s + \beta$ produces a valid X-Slits panoramic image.

The parameters of the camera's slits are completely defined by α and β. In fact, there is a simple qualitative relation: α determines the "depth" of the vertical slit and β its "horizontal" shift. Thus a stereo pair can be generated qualitatively by using any two linear sampling functions with the same slope, and no additional calibration information is required. Moreover, we note that linear column sampling can be implemented by slicing the space-time volume of images, which can be done efficiently by mosaicing.

To see the relation between α, β and the slits' parameters, consider the plane Π defined by the camera's path and optical axis. Let X denote the axis overlapping the camera path on this plane (to be called "horizontal" axis), and Z denote the orthogonal axis (to be called "depth"). One slit of the virtual camera overlaps the X axis, and we call it therefore the "horizontal" slit. Let (X_0, Z_0) denote the intersection of the vertical slit with plane Π. This point is fully determined by the column sampling function $s(t)$, and we can show that $X_0 \propto \beta$ and $Z_0 \propto \alpha$.

In order to generate a stereo pair of panoramas, we generate two X-Slits images with two different sampling functions. Let the first sampling function be $t(s) = \alpha s + \beta$ for some α, β, and let the virtual camera corresponding to the panoramic image generated by this function have its vertical axis going through (X_0, Z_0). We can now generate a second X-Slits image by using the sampling function $t'(s) = \alpha s + \beta + \delta$; the virtual camera corresponding to the panoramic image generated by this function has its vertical axis going through $(X_0 + \Delta, Z_0)$. Thus we have created two images with identical horizontal slit, and with two vertical slits shifted horizontally by Δ one with respect to the other. This is going to be our stereo pair, where the free parameter δ is tuned to allow for comfortable image fusion.

We tested our method on a sequence taken by a sideways moving pinhole camera; examples are shown in Fig. 5. We show a number of stereo pairs, obtained by moving the pair of vertical slits around. For comparison, we also show the stereo pair obtained from the same sequence by the "stereo panorama" algorithm.

A complete analysis of X-slits stereo and its epipolar geometry can be found in.[8] We will give some intuitive description of this stereo effect.

Examine for example the left slice in Fig. 4.b. This slice is a mosaic image generated by stitching together strips taken from the left side of the input images. A left strip in the input image is generated by rays coming from the right, and the entire mosaic image is built from rays as shown in Fig. 6.a. The right slice in Fig. 4.b is generated similarly by rays as described in Fig. 6.c. Since each slice is made from rays viewing the scene in a different direction, every pair is a stereo pair.

(a) (b)

Figure 5. Panoramic stereo pairs (Anaglyph image to be viewed in color in the electronic proceedings): (a) X-Slits (slices as in Fig. 4.c) and (b) pushbroom (slices as in Fig. 4.b). The images should be viewed in color, using cyan filter for the right eye and red filter for the left eye.

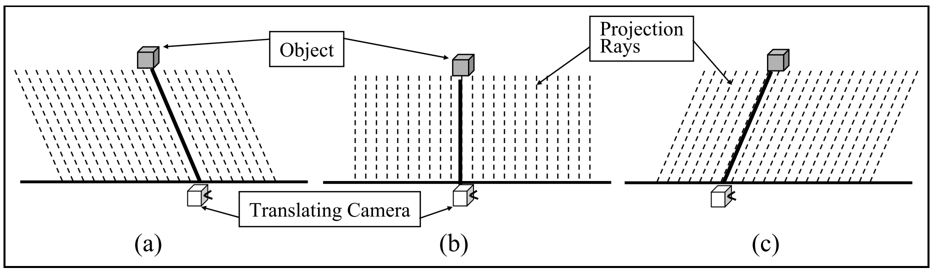

Figure 6. Panoramic images with directional rays. (a) Rays of a mosaic generated by strips from the right side of the input image, represented by the right slice of the volume in Fig. 4.b. (b) Rays of a mosaic generated by center strips, or the central slice in Fig. 4.b. (c) Rays of a mosaic generated by strips from the left, or the left slice in Fig. 4.b.

There is no limit to the number of different stereo views that can be generated, as the slices can be made as close to each other as we wish. This is an ideal situation for the generation of multiple stereo images for lenticular printing and for stereo displays that require multiple stereo views.

Diagonal parallel slices, as shown in Fig 4.c, also form a stereo pair. As described earlier, such diagonal slices are closer to perspective projection, and therefore are perceived as having less distortions.

5. EXPERIMENTAL RESULTS

New view generation is done by sampling columns from successive images and pasting them together into a stereo pair of X-Slits images. The parameters of the column sampling functions determine the location of the vertical slits of the stereo cameras. In the stereo sequences generated in our experiments, we manipulated the parameters of the sampling function so that the location of the vertical slit moved according to the desired ego-motion. Although the horizontal slit of the camera remained fixed, our results demonstrate a very compelling impression of stereo and camera motion.

5.1. Virtual Walkthrough

In Fig 7, we use an image sequence taken by a camera translating sideways to synthesize a new stereo sequence which corresponds to a camera moving forward.

Another example shows a panoramic stereo pair created by mosaicing from a sequence taken from a car driving in a parking lot (Fig. 8). A substantial parallax is visible between the two panoramic views.

Figure 7. The scene was filmed once by a sideways moving camera. A simulated forward motion shows a far-away stereo pair and a stereo pair from a closer location. The images were generated by mosaicing.

Figure 8. A panoramic stereo pair made by mosaicing an image sequence taken from a car driving in a parking lot.

6. CONCLUDING REMARKS

We presented a new non-perspective projection model, which is defined by two slits and a projection surface. Although this model can be physically realized, we propose to obtain appropriate images by mosaicing.

The application we described in this paper is image based rendering of stereo images. New view generation with the X-Slits projection is greatly simplified as compared with perspective new view generation, as it is performed by non-stationary mosaicing, or by slicing the space-time volume. The X-Slits theory helps the user to "drive" the slicing process in order to get the desired effect. When compared to traditional mosaicing, X-Slits images can be shown to be closer to perspective images than linear pushbroom images.

Using the proposed method we can also generate new images taken from "impossible" positions, like behind the back wall of a room or in front of a glass barrier. Movies with new ego motion can also be generated, such as forward-moving movies from a side-moving input sequence. Although not perspective, the movies generated in this way appear compelling and realistic.

Bibliography

[1] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall, "Mosaicing new views: The crossed-slits projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**, pp. 741–754, June 2003.

[2] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23**, pp. 279–290, March 2001.

[3] R. Hartley and R. Gupta, "Linear pushbroom cameras," in *Third European Conference on Computer Vision*, J. Eklundh, ed., pp. 555–566, Springer, (Stockholm, Sweden), May 1994.

[4] T. Pajdla, "Stereo with oblique cameras," *IJCV* **47**(1-3), pp. 161–170, 2002.

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

[6] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Second European Conference on Computer Vision*, G. Sandini, ed., pp. 237–252, Springer, (Santa Margherita, Italy), May 1992.

[7] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22**, pp. 1144–1154, October 2000.

[8] D. Feldman, T. Pajdla, and D. Weinshall, "On the epipolar geometry of the crossed-slits projection," in *9th International Conference on Computer Vision*, pp. 988–995, IEEE-CS, (Nice, France), October 2003.