

New view generation with a bi-centric camera

Abstract. We propose a novel method for new view generation from a rectified sequence of images. Our new images correspond to a new camera model, which we call a bi-centric camera; in this model the centers of horizontal and vertical projections lie in different locations on the camera's optical axis. This model reduces to the regular pinhole camera when the two projection centers coincide, and the pushbroom camera when one projection center lies at infinity. We first analyze the properties of this camera model. We then show how to generate new bi-centric views from vertical cuts in the epipolar volume of a rectified sequence. Every vertical cut generates a new bi-centric view, where the specific parameters of the cut determine the location of the projection centers. We discuss and demonstrate applications, including the generation of images where the virtual camera lies behind occluding surfaces (e.g., behind the back wall of a room), and in unreachable positions (e.g., in front of a glass window). Our final application is the generation of movies taken by a simulated forward moving camera, using as input a movie taken by a sideways moving camera.

Keywords: new view generation, image-based rendering, bicentric camera, pushbroom camera, epipolar volume

1 Introduction

Existing approaches to the problem of new view generation from images can be divided along the line of 3-D realism as follows:

- The goal of the first class of techniques is the generation of high quality and veridical new images. One class of techniques involves image generation by interpolation from dense samples [4, 8]. The basic idea is to initially build a lookup table using many images of the given scene taken from many different viewpoints. New views are then generated by resampling this (huge) lookup table. Another class of techniques achieves view interpolation and extrapolation via image transfer, using the epipolar geometry and structures such as the trifocal tensor [1]. See additional references in [7].
- The goal of the second class of techniques is the generation of new images for such purposes as compression and visualization. To this end the realistic quality of the image is sacrificed, emphasizing instead the generation of a “visually compelling” image. This class includes mosaicking techniques, which paste images together based on image similarity in overlapping areas [10]. Other techniques achieve view interpolation and extrapolation by learning the mapping between a set of stored images and various types of new images. Once again, see [7] for references.

Both classes of techniques can be very useful when used within the domain in which they are designed to operate. But these domains exclude many interesting cases. Ray sampling techniques require a lot of carefully collected data and huge memory. View transfer and learning techniques require image correspondence and often use or compute geometric constraints. Finally, mosaicking techniques generate images which do not typically deliver veridical sense of the three dimensional environment, and cannot be used reliably for purposes such as 3-D planning and reasoning.

Our approach to new view generation combines mosaicking and ray sampling methods, enjoying (we believe) the advantages of both. The new image is generated similarly to a “regular” panorama, by taking overlapping image strips from successive images. But, as in ray sampling methods, strips corresponding to different rays are sampled from each image in a systematic way. Like mosaicking, our method is an image based approach, which uses minimal assumptions and constraints about the input stream. But like ray-sampling, it is designed to generate geometrically veridical views that can be generated by a known camera model. Our new images are designed to give a geometrically consistent sense of the three-dimensional environment.

In Section 2, we present the mathematical properties of the projection model which describes the new views generated by our approach. It is not a true perspective camera model and therefore our new images cannot (even in principle) be taken by an ideal pinhole camera. Therefore the examples which we give should be used to judge whether the new images really give a compelling sense of three-dimensions.

In Section 3 we describe how to generate new bi-centric views. Specifically, we describe how to generate views taken from arbitrary new camera positions by simply taking different vertical cuts in the epipolar volume of the original sequence, and pasting the columns together. This simple setup requires that the input stream be generated by a camera translating along a line in 3-D, without changing its orientation or internal calibration (see discussion in Section 3).

In Section 4 we demonstrate a number of applications of our method: Starting from a rectified sequence taken by a camera moving sideways, we can generate images taken from virtual locations behind or in front of the moving camera. This allows us to generate large field of view images taken (virtually) from impossible locations, such as behind walls. Take, for example, an indoors scene; typically it is not possible to photograph a whole room even from the position at the back wall of the room, because the required field of view can be close to 180° . What we propose to do instead is: (i) first take a sequence of images by moving along the back wall of the room; (ii) then generate a bi-centric image taken from a virtual position which is *far behind* the back wall of the room. Our examples below show that the effect can be quite striking.

We can use the same principle in order to generate a simulated sequence where the simulated camera moves towards a scene, using as input a sequence generated by a camera moving sideways. This allows us to generate forward moving movies under conditions that do not normally make the generation of such movies possible; for example, when there are occluding surfaces in the scene (such as a wall behind the camera) or obstacles to motion (such as a glass window in front of the camera).

As our examples show, our simulated sequences are much more compelling than the alternative, which is the simulation of forward motion by zoom. When a camera zooms, all that happens is that the relative size of objects changes; the overall scene, however, remains the same - no occluded objects come to view. This is one reason why a zooming sequence does not give the viewer a comfortable feeling of forward motion. Our simulated sequences, on the other hand, give the impression of real forward motion, with occluded objects appearing and disappearing in accordance with the 3-D geometry.

Other applications of our method include: (i) the representation of a sequence by a bi-centric panorama instead of the pushbroom panorama, which is typically used when the camera is translating; (ii) new view generation for stereo viewing or 3-D TV.

The idea of sampling columns from images has been recently explored in [12, 11]. While [11] only used the constant sampling function (as in the traditional pushbroom panorama), [12] indeed used a varying sampling function for the columns. However, like the other ray-sampling methods described above, the proposed sampling method relied on the precise measurement of camera location and orientation along the way. The “vertical distortion” was considered an artifact that needed to be eliminated via the use of some domain knowledge or 3-D reconstruction. On the other hand, change of camera orientation was permitted.

2 The bi-centric camera model

A camera model is, in general, a mapping from \mathcal{R}^3 to \mathcal{R}^2 , describing which point in the real world ($\in \mathcal{R}^3$) is mapped to which point in the image ($\in \mathcal{R}^2$). In particular, a pinhole camera is such linear mapping when using homogenous (projective) coordinates to describe both world and image coordinate systems.

To describe the mapping of a projective camera, we use the following coordinate systems of the so-called “standard camera” [3] (see Fig. 1(a)). The world coordinates (X, Y, Z) of a point M in \mathcal{R}^3 are measured in the 3-D coordinate system whose origin is the focal center of the camera, X, Y denote the directions parallel to the horizontal and vertical image dimensions respectively, and Z denotes the direction of the optical axis of the camera (perpendicular to the image plane). The focal length of the camera, denoted by f , is the distance between the camera’s focal center and the image plane in this coordinate system. Finally, the image coordinates (x, y) of a point m in \mathcal{R}^2 are measured in the coordinate system whose origin is the intersection of the optical axis with the image plane, and x, y denote the directions of the horizontal and vertical image axes respectively.

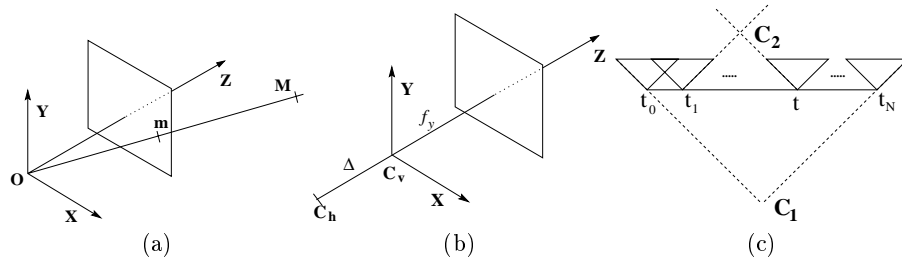


Fig. 1. (a) The standard camera. (b) Bi-centric camera. (c) A camera translating sideways; we also show virtual bicentric camera centers behind the scene (C_1) and in the scene (C_2).

Using this notation, an ideal calibrated pinhole camera can be described by the following mapping:

$$M = (X, Y, Z) \implies m = (x, y) = \left(f \frac{X}{Z}, f \frac{Y}{Z}\right) \quad (1)$$

To describe the calibrated bi-centric camera, we use the same 3-D coordinate system, see Fig. 1(b). The center of vertical projection in all the images C_v is fixed at the origin, but the center of horizontal projections C_h is located at another point $(0, 0, -\Delta)$ on the optical axis. This dissociation is the reason why we call our camera bi-centric.

The focal length in the vertical dimension is denoted by f_y and the focal length in the horizontal direction is $f_x = f_y + \Delta$. It follows that the bi-centric

camera model used in our method corresponds to the following mapping from \mathcal{R}^3 to \mathcal{R}^2 :

$$(X, Y, Z) \implies (x, y) = \left(f_x \frac{X}{Z + \Delta}, f_y \frac{Y}{Z} \right) \quad (2)$$

In general, an image created by a bi-centric camera cannot be generated by an ideal pinhole camera unless $\Delta = 0$. When $\Delta = \infty$, the bi-centric camera becomes a pushbroom camera, which corresponds to a vertical strip of sensors translating sideways [5]). In this case (2) reduces to

$$(X, Y, Z) \implies (x, y) = \left(X, f_y \frac{Y}{Z} \right)$$

One peculiarity of the bi-centric camera is the fact that straight lines are not mapped to straight lines (which is also the case for the pushbroom camera). More specifically, given any 3D line not perpendicular to the optical axis Z , denoted by $(a + c\lambda, b + d\lambda, \lambda)$, its projection is given by

$$(x, y) = \left(f_x \frac{a + c\lambda}{\lambda + \Delta}, f_y \frac{b + d\lambda}{\lambda} \right) \quad (3)$$

It follows from (3) that $\lambda = \frac{af_x - x\Delta}{x - cf_x}$ and $\lambda = \frac{bf_y}{y - df_y}$. Therefore, for every point (x, y) on the projection of the line

$$\frac{-\Delta}{f_x f_y} xy + \frac{d\Delta - b}{f_x} x + \frac{a}{f_y} y + (bc - ad) = 0$$

Thus the line is projected to hyperbolic segment. The boundaries of the projected hyperbolic segment correspond to the visible end-points of the line $Z = f_y$ (considering all points in front of the image plane) and $Z = \infty$. The image coordinates at these boundaries can be shown to be:

$$\begin{aligned} Z = f_y : & \quad x = a + cf_y, y = b + df_y \\ Z = \infty : & \quad x = cf_x, y = df_y \end{aligned} \quad (4)$$

Another interesting property of the bi-centric camera is the occlusion relations describing all the points which are projected to the same single point in the image. These relations determine the occlusion properties of this camera model: among all the points which satisfy a certain occlusion relation, the one closest to the camera occludes the others. In an ideal pinhole camera, an occlusion relation is an optical ray – a 3-D line which passes through the focal center of the camera.

We find the occlusion relations of the bi-centric camera by solving directly from (2) for all 3-D points which are projected to an image point (x, y) :

$$\left\{ \left(\frac{x}{f_x}(\lambda + \Delta), \frac{y}{f_y}\lambda, \lambda \right) \mid \lambda \in \mathcal{R} \right\}$$

This equation describes a ray (as in the models studied in [9, 11]). For all points with $x = x_0$, these rays intersect at $\left(\frac{x_0\Delta}{f_x}, 0, 0\right)$, whereas for all points with $y = y_0$, these rays intersect at $\left(0, \frac{-y_0\Delta}{f_y}, -\Delta\right)$. Note also that, rather than having a single intersection as in the pinhole model, all camera rays intersect one horizontal line $(\lambda, 0, 0) \forall \lambda$ and one vertical line $(0, \mu, -\Delta) \forall \mu$.

A bi-centric image contains distortions reminiscent of lens distortions (but otherwise very different). Let us look at the difference between the image coordinates of a 3-D point projected through a bi-centric camera to (x, y) , and the image coordinates of the same point projected through a pinhole camera to (x', y') . The pinhole camera is assumed to be identical to the bi-centric camera with $\Delta = 0$. From (2) it follows that:

$$x' = f_y \frac{X}{Z}, \quad x = f_x \frac{X}{Z + \Delta}, \quad y = y' = f_y \frac{Y}{Z}$$

and the bi-centric distortion is

$$\frac{x'}{x} = \frac{f_y \frac{X}{Z}}{f_x \frac{X}{Z + \Delta}} = \frac{f_y}{f_x} \left(1 + \frac{\Delta}{Z}\right) \tag{5}$$

It follows from (5) that the distortion of a bi-centric image depends on depth. There is no distortion near the image plane, for $Z = f_y$, because $x'/x = 1$. The distortion then increases as a function of $\frac{1}{Z}$ as the depth Z increases, for maximal distortion of $\frac{f_y}{f_x}$ for distant points.

It is interesting to compare bi-centric distortions to pushbroom distortions, where the depth-dependency of the distortion is maximal. Imagine a pushbroom camera recording a scene which contains the full moon at its center. The moon, being very far away, will be seen in the center of each image as the camera sweeps sideways; in the final pushbroom image, the image of the full moon will be a strip of light crossing the image horizontally from left to right. In the bi-centric image, on the other hand, the full moon will be projected to an ellipse with aspect ratio of f_x/f_y .

3 New view generation and the epipolar volume

In this section we describe a method to generate new bi-centric views from a rectified sequence of images. We initially assume (Section 3.1) that the camera is internally calibrated, and that it translates with constant speed in a direction parallel to the horizontal axis of its image plane. The uncalibrated case is addressed in Section 3.2.

For a camera translating along a line in 3-D space, we can construct the so-called ‘‘epipolar volume’’, i.e., the box obtained by stacking the images one in front of the other. Our key result is that every planar cut of this volume, which is parallel to the Y axis of the box (or the vertical axis of the images), corresponds to a new view of the scene taken with a bi-centric camera.

3.1 An internally calibrated camera

Assume that the input is a sequence of images, taken by an internally calibrated pinhole camera whose projection model is given in (1). Assume further that the camera is translating with constant speed sideways from left to right, in a direction parallel to the horizontal axis of its image plane. A schematic illustration of this setup is given in Fig. 1(c), showing a top-down view of the sequence of positions covered by the real camera's focal center, and their corresponding fields of view.

Suppose now that we generate a new image by taking successive vertical strips from successive images (as is often done in mosaicking algorithms). However, instead of taking the central strip from each image, we do something else: As the first strip in our new image, we take the left-most strip from the first image in the sequence. As the last strip in our new image, we take the right-most strip from the last image. In between, we incrementally sample successive columns of the successive images and paste them side by side. The result is a bi-centric image, whose horizontal size is larger than its vertical size.

Fig. 2(a) illustrates how this image can be generated by a single virtual bi-centric camera. More specifically, suppose we have a sequence of images taken with a pinhole camera with focal length $f = f_y$ at different positions along a segment of the X axis. The projection model of the pinhole camera positioned at $(X_0, 0, 0)$ is given by $(x, y) = (f \frac{X-X_0}{Z}, f \frac{Y}{Z})$. We use $[-r, r]$ to denote the horizontal range of each pinhole image, and $X \in [-l, l]$ to denote the range of positions of the pinhole camera centers in R^3 .

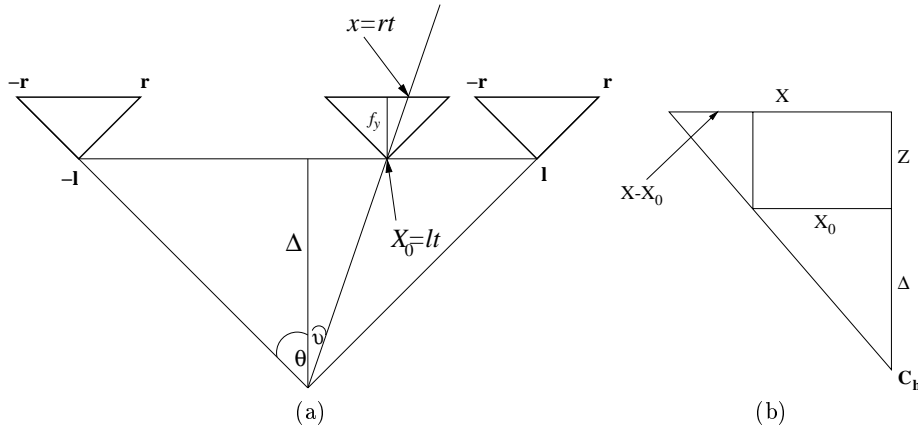


Fig. 2. (a) Bi-centric image formation; (b) bi-centric projection into the X - Z plane.

The horizontal range of the new image is $[-(r+l), (r+l)]$. For each $t \in [-1, 1]$, we define the $(l+r)t$ column of the new image to be the rt column of the pinhole camera positioned at $(lt, 0, 0)$ ($lt = X_0$ above). Now, for each column

$x \in [-(r+l), (r+l)]$ in the new image, $t = \frac{x}{l+r}$ and therefore it follows from $f \frac{X-lt}{Z} = rt$ that

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

where

$$f_x = f \frac{l+r}{r}, \quad f_y = f, \quad \Delta = f \frac{l}{r}$$

The above, together with the invariable property of the pinhole cameras $y = f_y \frac{Y}{Z}$, yield the same model as described by (2).

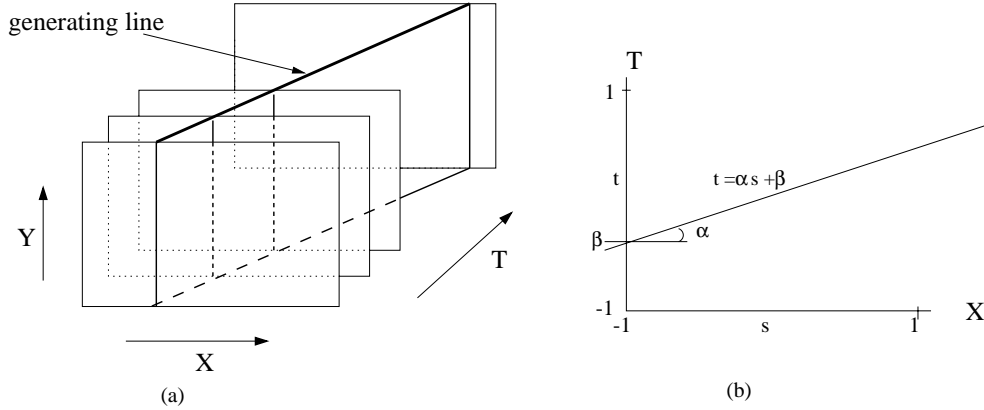


Fig. 3. The epipolar volume: (a) an oblique view and an example vertical cut, (b) top view.

In order to generalize this scheme, we first need to define the epipolar volume (see Fig. 3). It is the volume obtained by stacking the images one in front of the other; its axes are the horizontal image axis X , the vertical image axis Y , and time T . A vertical planar cut in this volume is defined by a generating line in the $X - T$ plane. Every such cut defines a collection of successive vertical image strips, which are combined to generate a new image - the *vertical cut image*. Assuming dense epipolar volume, we state the following result:

Result: every *vertical cut image*, defined by a vertical cut in the epipolar volume, provides a bi-centric image.

In order to understand this result, consider the 2-D graph where one axis corresponds to the columns in each image, the second axis corresponds to all positions of a pinhole camera along the X axis (or equivalently, the $X - T$ plane of the epipolar volume). Let $s \in [-1, 1]$ denote the index of the column axis (X), and t denote the index of the motion axis (T), corresponding to all pinhole

camera positions $(lt, 0, 0)$ for some $l \in \mathcal{R}$. If the pinhole cameras have a field of view of 2θ , then $s = \frac{\tan \nu}{\tan \theta}$ for the bi-centric column at the location X_0 and angle ν from the optical axis (see Fig. 2(a)).

Let us consider now the line in s - t space given by $t = \alpha s + \beta$ (see Fig. 3(b)). For each s , the points projected by the camera located at $(lt, 0, 0)$ onto that camera's s column come in 3-D from the line (see Fig. 2(a)):

$$X = l(\alpha s + \beta) + \lambda s f_y \tan \theta, \quad Y = \lambda y, \quad Z = \lambda f_y \quad \forall \lambda$$

This 3-D line is projected onto point $(s f_x \tan \theta, y)$ in the vertical cut image; therefore

$$x = f_x \tan \theta s = f_x \tan \theta \frac{X - \beta l}{\alpha l + \lambda f_y \tan \theta} = f_x \tan \theta \frac{X - \beta l}{\alpha l + Z \tan \theta}$$

as well as $y = f_y \frac{Y}{Z}$. Setting $X_0 = \beta l$ and $\Delta = \alpha l / \tan \theta$, we obtain

$$x = f_x \frac{X - X_0}{Z + \Delta}$$

Finally, we set $f_x = f_y + \Delta$.

In summary, by sampling the epipolar volume along a vertical cut defined by $t = \alpha s + \beta$, the center of horizontal projections is located at

$$(X, Z) = \left(\beta l, \frac{-\alpha l}{\tan \theta} \right) \quad (6)$$

Vice versa, from (6) it follows that in order to move the horizontal projection center to $X = X_0$, $Z = Z_0$, we must sample the vertical cut of the epipolar volume defined by $t = \alpha s + \beta$, where $\beta = \frac{X_0}{l}$ and $\alpha = -\frac{Z_0}{l} \tan \theta$.

Note that if the slope of the vertical cut is positive ($\alpha > 0$), the horizontal projection center of the camera is located backwards behind the moving camera. In this case (see Fig. 1(c)), the leftmost column of the bicentric image comes from the leftmost column of one of the input images. The rightmost column of the bicentric image comes from the rightmost column of another input image that occurs later in the sequence. On the other hand, if the slope is negative ($\alpha < 0$), the horizontal projection center of the camera is located forwards, *inside* the scene. In this case, the leftmost column of the bicentric image comes from the *rightmost* column of one of the input images. The rightmost column of the bicentric image comes from the *leftmost* column of a later input image. Consequently, to generate consistent forward moving sequence, the inside-the-scene images have to be mirrored.

The construction above gives us a bi-centric image where the virtual image plane is located at $Z = f_y$. The analysis of the bi-centric depth-dependent distortion in Section 2 tells us the following: if we photograph a fronto-parallel square at depth Z , it will appear in the image as a rectangle with aspect ratio of $\frac{f_y}{f_x} \left(1 + \frac{\Delta}{Z}\right)$. Only squares on the image plane $Z = f_y$ will appear as squares. This is not an optimal choice for best “visual appeal”. Thus we normalize the images in the last step of the new view construction as follows:

- We choose a fixed depth Z_0 and normalize the images by the factor $\frac{f_y}{f_x} \left(1 + \frac{\Delta}{Z_0}\right)$. Note that this factor depends on Δ and is thus different for each new image.
- As a result, planar objects at depth Z_0 will appear in the image with veridical aspect ratio.
- Planar objects at depth $Z > Z_0$ will appear horizontally elongated, while planar objects at depth $Z < Z_0$ will appear vertically elongated. In either case, the distortion is proportional to $|Z - Z_0|$.

We note that before normalization, as we move the horizontal projection center, the image of a scene point moves horizontally on the image plane towards the line $x = 0$. Thus the optical flow is linear and horizontal, and the FOE is a vertical line (instead of a point). The final image scaling adds a vertical component to the optical flow, as a result of which the (combined) 2-D motion field is not linear.

3.2 The uncalibrated camera

Suppose the camera is not internally calibrated, but its internal calibration remains fixed throughout the sequence. We can still generate an image using the technique from the previous section. Below we show that the resulting image is still bi-centric and we show the effects of the internal calibration on the image.

Let (X, Y, Z) denote the 3-D coordinates of a point, and (x_0, y_0) denote the image coordinates of the point in the pinhole camera C_0 located at $(X_0, 0, 0)$. In the uncalibrated camera C_0 , the image formation equations from (1) become the following:

$$\begin{aligned} (X, Y, Z) &\implies (p_0, q_0) = \left(\frac{X - X_0}{Z}, \frac{Y}{Z}\right) \\ &\implies (x_0, y_0) = (f_y p_0 + s q_0 + u, f_y q_0 + v) \end{aligned} \tag{7}$$

The image coordinates in the “standard” coordinate system of C_0 are denoted by (p_0, q_0) and are only used in an intermediate step. The second mapping from (p_0, q_0) to (x_0, y_0) is determined by the internal calibration parameters of the camera. The expressions above, when written in a 2-D homogeneous coordinate system, reflect the multiplication of the image by the following internal calibration matrix:

$$\mathbf{K} = \begin{bmatrix} f_y & s & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

For clarity and simplicity of notation we assume the same focal length for both horizontal and vertical axes in the original uncalibrated pinhole cameras. The generalization of the proof to the general case is straightforward.

Let (x, y) denote the image coordinates of the point (X, Y, Z) in the final bi-centric image. In the original, C_0 , camera $x_0 = rt$, while in the bi-centric

image $x = (r + l)t$. Since $(r + l)/r = f_x/f_y$ (see Fig. 2(a)), we have $x = x_0 f_x/f_y$ and $y = y_0$. Using (7), we get:

$$x = f_x \cdot p_0 + s \frac{f_x}{f_y} \cdot q_0 + \frac{f_x}{f_y} u, \quad y = f_y \cdot q_0 + v$$

Let (p, q) denote the equivalent of standard image coordinates for (X, Y, Z) in the bi-centric camera:

$$(X, Y, Z) \implies (p, q) = \left(\frac{X}{Z + \Delta}, \frac{Y}{Z} \right)$$

We now use this definition and (7), as well as the equality $\frac{X - X_0}{X} = \frac{Z}{Z + \Delta}$ (see Fig. 2(b)), in order to obtain the relation between (x, y) and (p, q) :

$$\begin{aligned} x &= f_x \cdot \frac{X - X_0}{Z} + s \frac{f_x}{f_y} \cdot \frac{Y}{Z} + \frac{f_x}{f_y} u = f_x \cdot \frac{X}{Z + \Delta} + s \frac{f_x}{f_y} \cdot \frac{Y}{Z} + \frac{f_x}{f_y} u \\ &= f_x \cdot p + s \frac{f_x}{f_y} \cdot q + \frac{f_x}{f_y} u \\ y &= f_y \cdot \frac{Y}{Z} + v = f_y \cdot q + v \end{aligned} \tag{9}$$

We therefore conclude that, if we construct a new image from a vertical cut in the epipolar volume, and if the volume is generated by a camera whose internal calibration is matrix K from (8), then the resulting image is still bi-centric with the following effective internal calibration matrix:

$$\mathbf{K}_{bc} = \begin{bmatrix} f_x & s \frac{f_x}{f_y} & \frac{f_x}{f_y} u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{f_x}{f_y} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} K$$

4 Results

We used a simulated 3-D scene and two real sequences to study bi-centric images and demonstrate the applications discussed above.

4.1 Simulations

The simulated scene includes a set of boxes, whose organization on the plane forms a grid, surrounded by a wall. A sequence was generated by the simulation of a pinhole camera moving sideways behind the boxes but inside the wall. Then virtual bicentric images were generated at different locations, including behind the wall. We compared virtual bicentric images (Fig. 4(a)) with the ideal perspective images taken from the location of the new horizontal projection center with and without occluding surfaces (Fig. 4(b),(c)). We also compared with panoramic images generated by traditional methods - pushbroom and panning panorama (Fig. 4(d),(e)).

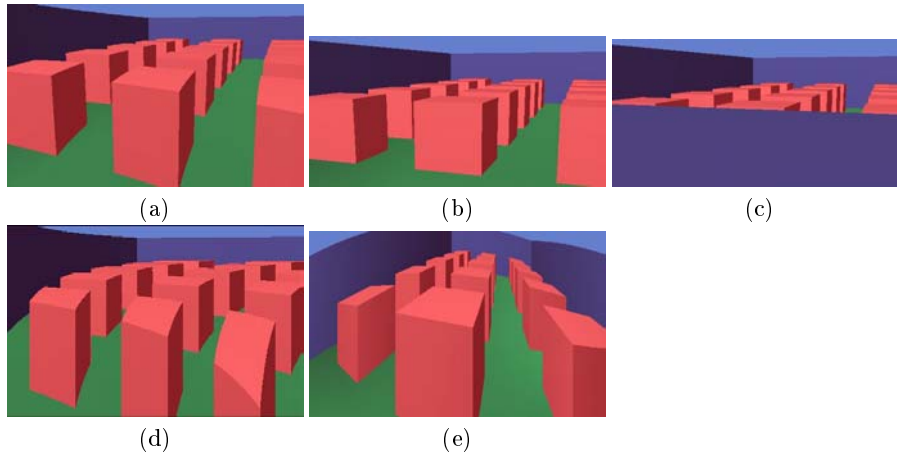


Fig. 4. Simulation results: a virtual bicentric image (a) is compared with the corresponding pinhole image as it would have looked without (b) and with (c) occluding surfaces. We also show two traditional panoramic images generated from roughly the same position as (a): (d) pushbroom, (e) cylindrical projection (panning, or self-rotating, camera).

4.2 Real sequences

We took two real sequences using a camera with 48° field of view. We moved the camera on a straight track, with the image plane roughly parallel to the direction of motion, at approximately constant speed. Fig. 5 presents three frames of each sequence. For the cafeteria scene, 208 frames were captured on a 4.4m track. For the laboratory scene, 360 frames were captured on a 2m track. The bicentric images generated by our method are shown in Fig. 6 (a) and (b). For comparison, we used the same camera to take an image with roughly the same field of view in the laboratory (Fig. 6(c)).

We also generated a dolly (move in) and a tracking (move in and sideway) sequence for each scene, by picking subsets of consecutive frames in the input sequence to produce a series of bicentric images. The sequences start with the images shown in Fig. 6(a) and (b) respectively. Fig. 7 shows the middle and last frames of the output sequences of the cafeteria scene. We scaled the height of the last frame so that the aspect ratio of the windows at the back maintained its veridical value. We interpolated the height for the rest of the frames. As another application, a stereo pair of the cafeteria scene is shown in Fig. 7(e).

Similarly, Fig. 8 shows the middle and last frames of the two output sequences of the laboratory scene. In this case, the height of the last frame is chosen to maintain the veridical aspect ratio of the column in the middle of the scene.

Lastly, for comparison, we generated panoramas using the pushbroom camera model. Figs. 9(a),(b) were generated by sampling the middle column of each frame in the input sequences. For the cafeteria scene, the panorama was scaled

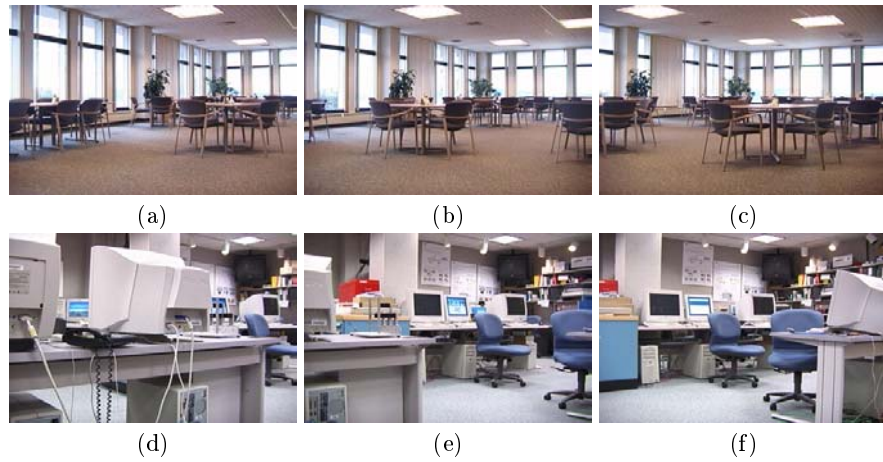


Fig. 5. A few frames (360x240) of the input sequences. In the cafeteria scene: (a) first frame; (b) middle frame; and (c) last frame. In the laboratory scene: (a) first frame; (b) middle frame; and (c) last frame.

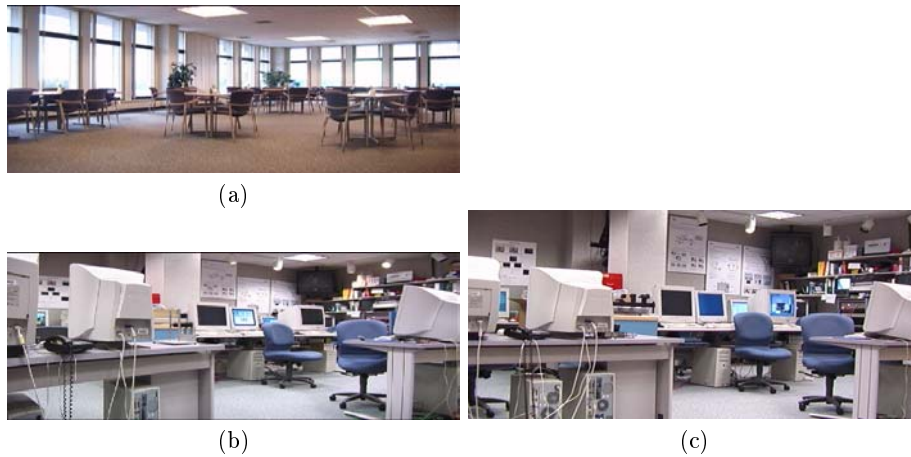


Fig. 6. Bicentric images generated from the sequences: (a) cafeteria scene; (b) laboratory scene. (c) Same scene in the laboratory captured by a projective camera at approx. 2m back from where the sequence was taken.

such that the aspect-ratio of the table and chairs in the middle is veridical. For the laboratory scene, the scale was fixed such that the computer, the table and the monitor on the left side of the image have the correct aspect-ratio.

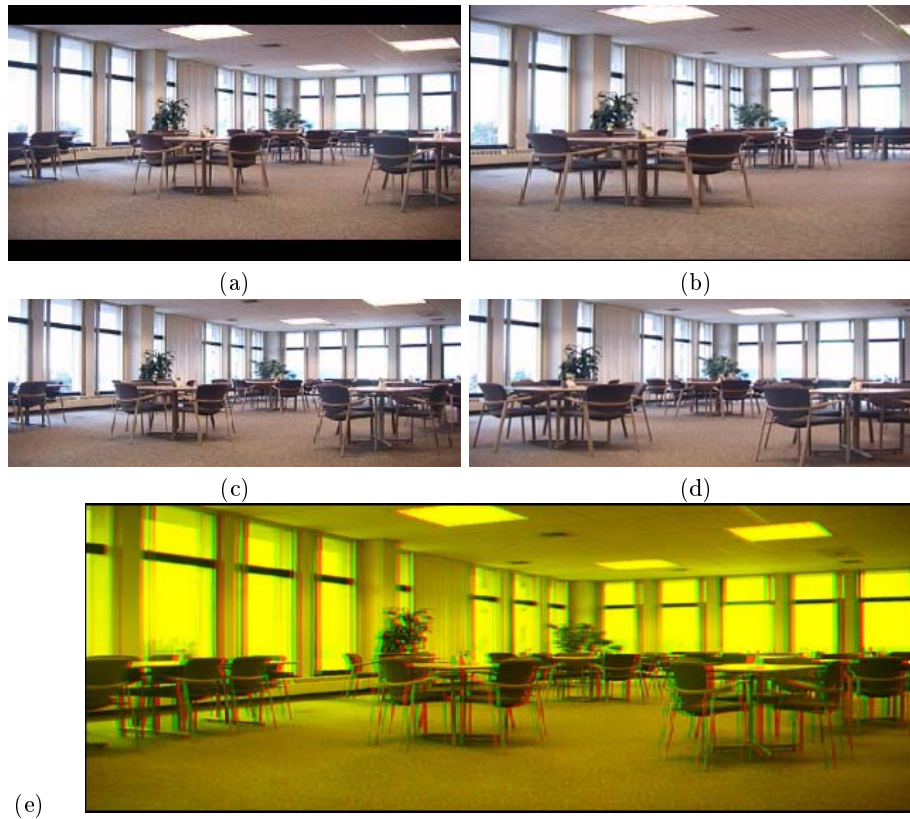


Fig. 7. Bicentric sequences of the cafeteria scene: (a) Middle frame of the dolly sequence (260 frames), generated from frames 70-137 of the input sequence; (b) Last frame of the dolly sequence. The virtual camera center is *in-the-scene*. This frame is generated from frames 77-130 of the input sequence, with leftmost column from the leftmost column of frame 130 and rightmost column from the rightmost column of frame 77. (c) Middle frame of the tracking sequence (208 frames), generated from frames 78-182; (d) Last frame of the tracking sequence, generated from frames 52-53. (e) A stereo pair - two bi-centric images obtained from two nearby positions; to view, use red-green stereo glasses and use the red on the left eye. The sequences can be obtained from the anonymous website <http://www.geocities.com/bicentriccamera/> as: dolly sequence - cafe-dolly.mpg, tracking sequence - cafe-tracking.mpg, stereo sequence - stereo.zip.

5 Conclusions and future work

We presented a novel method for new view generation from rectified sequences of images. Our method borrows from mosaicking and ray sampling approaches, enjoying the advantages of both. As in regular mosaicking, a new image is created by overlapping image strips from consecutive input images. However, as in

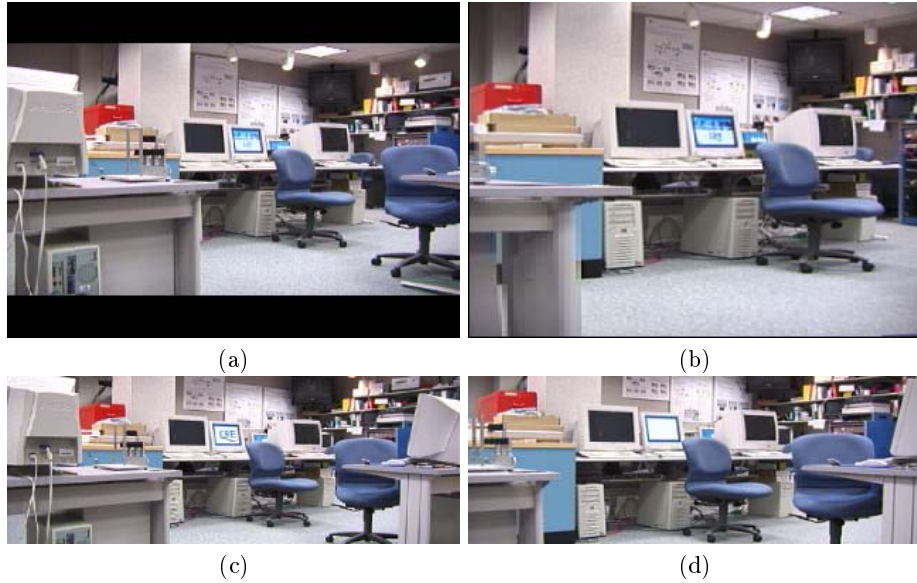


Fig. 8. Bicentric sequences of the laboratory scene: (a) Middle frame of the dolly sequence (300 frames), generated from frames 135-225 of the input sequence; (b) Last frame of the dolly sequence. The virtual camera center is *in-the-scene*. This frame was generated from frames 90-270 of the input sequence, with leftmost column from the leftmost column of frame 270 and rightmost column from the rightmost column of frame 90. (c) Middle frame of the tracking sequence (360 frames), generated from frames 117-297; (d) Last frame of the tracking sequence, generated from frames 53-54. Once again, the sequences can be obtained from the same anonymous website as: dolly sequence - lab-dolly.mpg, tracking sequence - lab-tracking.mpg.

ray sampling methods, strips are sampled from each image in a consistent and systematic manner. Newly created images correspond to a new camera model that we call a bi-centric camera; in this model the centers of horizontal and vertical projections lie in different positions on the camera optical axis. We have analyzed the properties of this model and demonstrated several applications, including the generation of panoramic images, the generation of images where the virtual camera lies in unreachable positions, and finally the generation of movies taken by a simulated forward moving camera using as input a sequence obtained by a sideways moving camera.

While our present approach is constrained by the requirement that the input sequence is obtained by a camera which translates sideways in constant speed, our ultimate goal is to generate new views and new movies from an arbitrary sequence of images. One solution involves the registration of the original images with each other using the homography of the plane at infinity. This computation requires either internal calibration, or some domain knowledge such as parallel

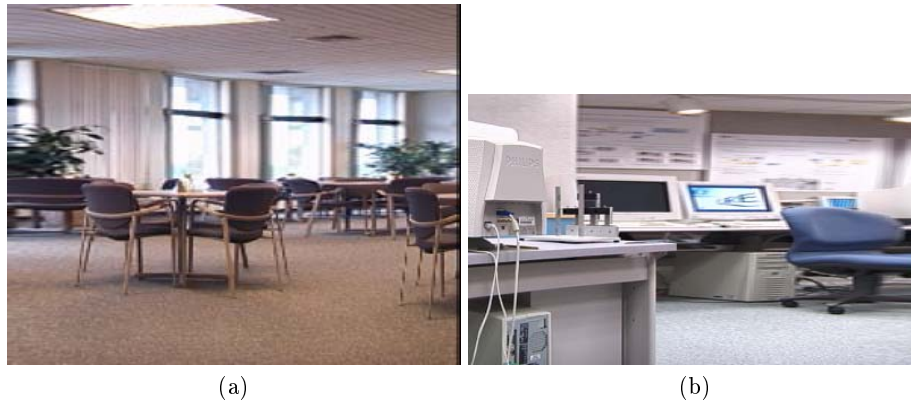


Fig. 9. (a) Pushbroom panorama of the cafeteria scene. (b) Pushbroom panorama of the laboratory scene.

lines in the scene. Our current efforts focus on the elimination of jitter that might be added to the motion of the camera, and/or deviation from constant speed.

References

1. S. Avidan, and A. Shashua. *Novel view synthesis in tensor space*. in *Proc. CVPR*, pp. 1034–1040, June 1997.
2. A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *Proc. ECCV*, pp. 846–861, Springer-Verlag, June 1998.
3. O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, MA, 1993.
4. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. *The lumigraph*. In *Proc. SIGGRAPH'96*, pp. 43–54, Aug 1996.
5. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.
6. M. Irani, P. Anandan and D. Weinshall. From reference frames to reference planes: Multiview parallax geometry and applications. In *Proc. ECCV*, pp. 829–845, 1998.
7. S. B. Kang. A survey of image-based rendering techniques. In *Videometric VI*, vol. 3641, pp. 2-16, Jan 1999.
8. M. Levoy, and P. Hanrahan. *Light field rendering*. In *Proc. SIGGRAPH'96*, pp. 31–42, Aug 1996
9. Tomáš Pajdla. Stereo with oblique cameras. TR CTU–CMP–2001–32, K333 FEE Czech Technical University, Prague, Czech Republic, 2001.
10. S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *Proc. CVPR*, pp. 338–343, June 1997.
11. S. M. Seitz, *The Space of All Stereo Images*. In *Proc. ICCV*, 2001.
12. T. Takahashi, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Arbitrary view position and direction rendering for large-scale scenes. In *Proc. CVPR*, pp. 296–303, June 2000.