Eurographics Symposium on Rendering (2007)
Jan Kautz and Sumanta Pattanaik (Editors)

General Linear Cameras with Finite Aperture

Andrew Adams and Marc Levoy

Stanford University

Abstract

A pinhole camera selects a two-dimensional set of rays from the four-dimensional light field. Pinhole cameras are a type of general linear camera, defined as planar 2D slices of the 4D light field. Cameras with finite apertures can be considered as the summation of a collection of pinhole cameras. In the limit they evaluate a two-dimensional integral of the four-dimensional light field. Hence a general linear camera with finite aperture factors the 4D light field into two integrated dimensions and two imaged dimensions. We present a simple framework for representing these slices and integral projections, based on certain eigenspaces in a two-plane parameterization of the light field. Our framework allows for easy analysis of focus and perspective, and it demonstrates their dual nature. Using our framework, we present analogous taxonomies of perspective and focus, placing within them the familiar perspective, orthographic, cross-slit, and bilinear cameras; astigmatic and anastigmatic focus; and several other varieties of perspective and focus.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation

1. Introduction

Point perspective has been well understood since the Renaissance, and is the geometric foundation of much of computer graphics. On occasion, in technical drawings especially, the point is moved to infinity and an orthographic projection is used. Recently, however, new kinds of perspective have been introduced, in which rays do not intersect at a point, even at infinity.

Gupta and Hartley [GH97] describe the *pushbroom* camera, motivated by the geometry of satellite imagery. Zomet et al. [ZFPW02] generalize this notion to the *cross slit* camera, which selects a family of rays passing through two lines in space. Pajdla [Paj02] describes *oblique* cameras, in which no two rays intersect (linear oblique cameras are also known as *bilinear* cameras). The work of Yu and McMillan [YM04b] then collected and generalized these cameras, classifying them as two-dimensional slices of the four dimensional space of rays passing between two planes.

These models all assume the generalized equivalent of a pinhole camera - each pixel records a single ray. Real cameras, by contrast, integrate over a finite aperture, forming a projected integral of ray space as opposed to just a slice. We present a mathematical model for general linear cameras which first simplifies the work of Yu and McMillan, and then extends it to cover focus. In our model, focus and perspective can be characterized by the eigenspaces of certain 2x2 matrices.

While this framework is primarily a theoretical contribution, it has several potential uses. First, it can be used for optical design. Specifically, unlike ray transfer matrices [Hal64], this framework can be used to characterize optical systems with cylindrical components, such as those used in some motion capture systems. Our framework can also be used as a tool to understand recent multiperspective techniques (such as [RGL04]). The most immediate utility of this framework is to render novel views from light fields or geometry.

Implementation of the viewer will be discussed in section 6. First, we introduce our parameterization (below), construct taxonomies of general linear perspective (section 3) and general linear focus (section 4), and discuss how they can describe general integral projections of ray space (section 5).

2. Parameterization of the Light Field

In this work we will parameterize a ray using four spatial coordinates. In figure 1, the ray $((u,v),(x,y))$ passes through the two points $(u,v,0)$ and $(x,y,1)$. This two-plane parameterization (described in [LH96]) exhibits a symmetry between space and angle, which we capitalize upon. The scene

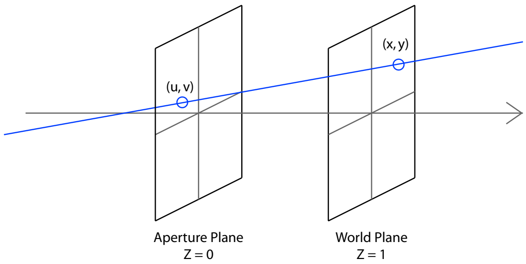

Figure 1: *A ray $((u,v),(x,y))$ in the two plane coordinate system.*

under consideration will typically be near the plane $z=1$, so the x and y coordinates of a ray denote a point in the scene, while the u and v coordinates describe the angle from which it is viewed. The plane $z=1$ will be called the world plane, while the plane $z=0$ will be called the aperture plane.

Our coordinate system is defined by the placement of these two planes. Hence they are parallel by definition. However, they need not be parallel in their containing coordinate system. They will intersect along some line, and all interesting events in this framework will occur on planes intersecting the same line. The transform of space implied by the coordinate system of the planes maps this line to infinity. In other words, the Scheimpflug condition holds (this fact is used in this context in [VGT*05]).

Within this two-plane parameterization of the light field we now describe and construct a taxonomy of general linear perspective.

3. Characterizing General Linear Perspective

The two most salient features of a perspective view are the location of the eye and the view direction. In our model the view direction is determined by the placement of the planes, with respect to which our viewing direction is along the positive Z-axis. The location of the eye is fixed by the meeting point of the rays that form our view.

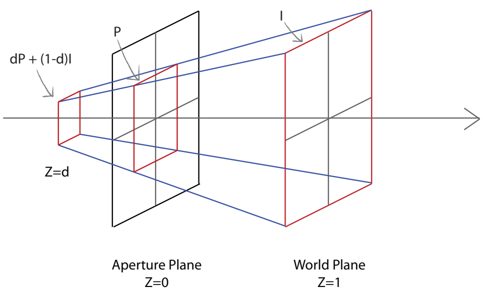

Figure 2: *The cross section of a bundle of rays with (u,v) coordinates equal to $P(x,y)$ can be determined at arbitrary depth by linear interpolation between the value at $z=0$, P, and the value at $z=1$, I.*

Consider the bundle of rays B_{xy} with ray space coordinates $(P(x,y),(x,y))$, where $(x,y)\in[-1,1]^2$, and P is some 2x2 matrix. B_{xy} is a 2D linear slice of 4D ray space, and hence forms a linear camera. This model covers all of the general linear cameras of Yu and McMillan [YM04b] except for the epipolar camera, discussed in section 5. For example, when P is zero we have a simple point perspective camera ($B_{xy}=((0,0),(x,y))$), and when $P=I$, we have an orthographic camera ($B_{xy}=((x,y),(x,y))$).

At the world plane ($z=1$) the cross section of B_{xy} describes a square. At the aperture plane ($z=0$) the cross section is a square transformed by P. Rays are linear, so we can linearly interpolate or extrapolate to determine the cross-section of the bundle at an arbitrary depth (Figure 2). If at $z=1$ the transformation of the square is the identity I, and at $z=0$ the transformation is P, then at $z=d$ the transformation is the 2x2 matrix:

$$P'_d=(1-d)P+dI \tag{1}$$

The rays in B_{xy} meet at a point when P' is rank zero, and they meet along a line when it is rank one. The kernel of a matrix is invariant under scaling, so if we parameterize depth by $\lambda=\frac{d}{d-1}$ then we can equivalently use the kernel of:

$$P''_\lambda=P-\lambda I \tag{2}$$

This kernel is described by the eigenvalues and eigenvectors of P. If a perspective view has an eigenvector v with real eigenvalue α, then the rays all intersect a line. This line is at depth $z=\frac{\alpha}{\alpha-1}$, is parallel to the world and aperture planes, and is normal to v, as P'_z zeroes that direction. Two distinct eigenvectors with the same real eigenvalue will bring the rays to a point at the corresponding depth. There are a very limited number of possibilities for the eigenspaces of a 2x2 matrix. The eigenvalues are the roots of the characteristic quadratic of the matrix. They may be real and equal, real and distinct, or complex conjugates. If the eigenvalues are distinct, then they must each be associated with a one-dimensional eigenspace (and hence an eigenvector). If they are equal, then they are associated with an eigenspace that is either one-dimensional (deficient), or two-dimensional.

Point Perspective Cameras. 2x2 matrices with equal eigenvalues and a two-dimensional eigenspace are all scales of the identity matrix. All point perspective cameras therefore have P matrix equal to λI. As shown in the top row of figure 3, when $\lambda<0$, the point of view is between the world and aperture planes. When $0<\lambda<1$ the point of view is behind the world plane. As λ approaches 1, the point of view tends towards negative infinity, and hence an orthographic camera. In general, the point of view is on the z axis at $\frac{\lambda}{\lambda-1}$.

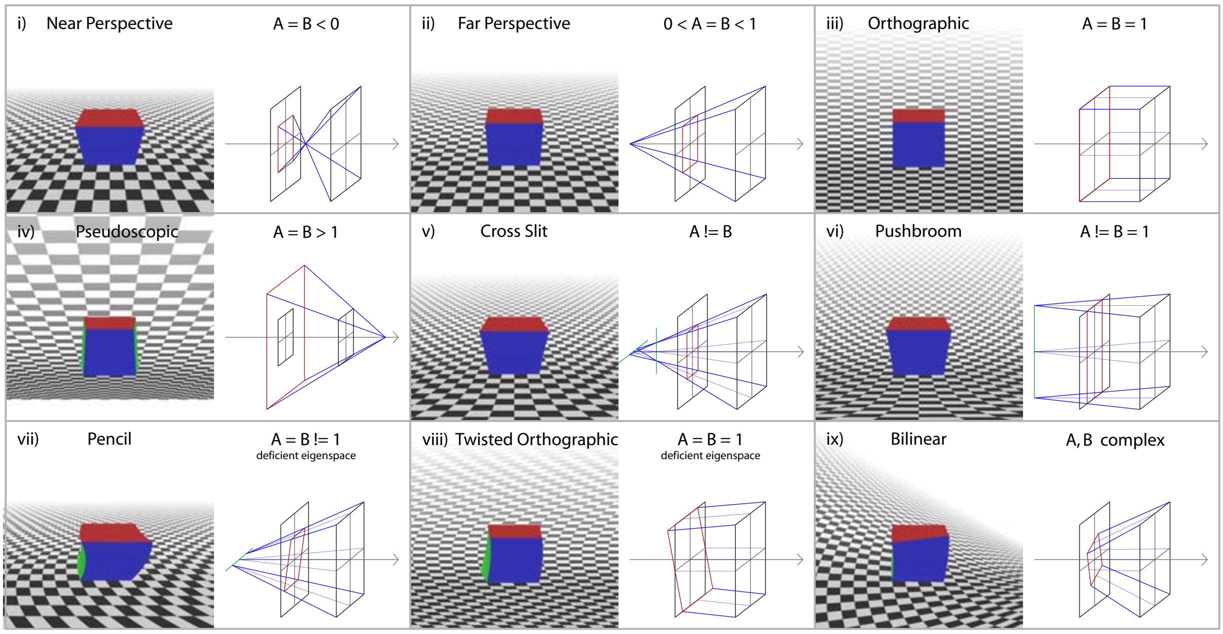

Figure 3: *Different conditions on the eigenvalues A,B of the matrix P give rise to different kinds of linear cameras. On the right in each diagram we see the rays used to construct the image on the left, of a cube on a checkered plane. P is represented as the red parallelogram.*

Setting $\lambda > 1$ produces a center of perspective beyond the world plane - a pseudoscopic view (Figure 3.iv). There are two ways to think of this type of camera. One can consider it a view from the aperture plane along a converging set of rays. This explains why four sides of the cube are visible, and why the squares on the checkerboard get larger as you go up the image. The other way to understand a pseudoscopic image is to treat the view as a conventional perspective from beyond the world plane looking back towards the aperture plane, but with an inverted depth buffer test.

Cross Slit Cameras. Matrices with distinct real eigenvalues produce cross slit cameras [ZFPW02] (Figure 3.v). An eigenvalue α with eigenvector v indicates a slit at $z = \frac{\alpha}{\alpha-1}$ in the direction normal to v. Point perspective cameras are as a special case of these, for which the slits are unaligned and at the same depth. The cross slit camera is of the two truly general linear cameras - the family has four degrees of freedom, and a small perturbation in its P matrix will still result in a cross slit camera. If one of the slits of a cross slit camera is at infinity, we obtain a pushbroom camera [GH97], shown in figure 3.vi.

Pencil Cameras. Matrices with equal eigenvalues but a deficient eigenspace produce pencil cameras (Figure 3.vii). As with all linear cameras, pencil cameras are defined by two linear constraints on ray space. One linear constraint is that all rays must pass through a particular slit, leaving a 3D family of possible rays. The second constraint selects allowable incoming angles of rays to the slit. The orthographic version of the pencil camera, for which the eigenvalues are one, is known as the twisted orthographic camera, and is shown in figure 3.viii. A pencil camera may occur in practice by horizontally translating a camera that uses rolling shutter, or equivalently photographing a moving object with a rolling shutter camera.

Bilinear Cameras. The bilinear camera (Figure 3.ix) is the other truly general linear camera. The eigenvalues of its matrix are complex conjugates. No two rays imaged by this camera intersect anywhere in space. Bilinear cameras can come arbitrarily close to point perspective cameras. For example, if P is a scaled very slight rotation then the family of rays will neck down almost to a point.

Relationships Between Cameras. The eigenvalues of P are the roots of its characteristic quadratic. If the discriminant of this quadratic is negative, the eigenvalues are complex, if it is positive, they are real and distinct. If the discriminant is zero, the eigenvalues are equal. Therefore, in the 4D space of P matrices, the pencil cameras form a 3D manifold separating the bilinear cameras from the cross slit cameras. The point perspective cameras are a 1D subset of the pencil cameras. All cameras, except those which are purely bilinear, have an orthographic variant. The 2D set of pushbroom cameras intersects the 3D manifold of pencil cameras at the 1D family of twisted orthographic cameras. This set intersects the point perspective cameras at the orthographic camera. These relationships are illustrated on the left in figure 4.

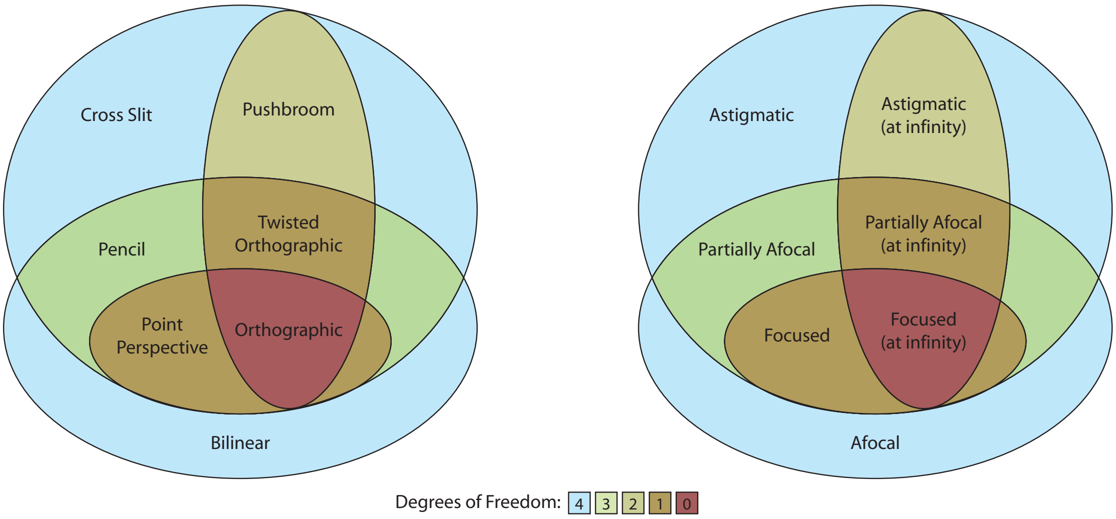

Figure 4: *Parallel Venn diagrams illustrating the set relationships between the various linear cameras on the left, and the various types of focus on the right. Colors indicate the dimensionality of each subset.*

4. Characterizing General Linear Focus

The most salient feature of a finite aperture view of a scene is the depth of the plane of best focus. For such a view one can define the output image *Im* as a linear projected integral over ray space. The domain of integration Ω will define the shape of the out of focus blur, or bokeh.

$$Im(x,y) = \int_{\Omega} L((u,v)+P(x,y),(x,y)+F(u,v))dudv \quad (3)$$

At output pixel (x,y) our model integrates the bundle of rays $((u,v)+P(x,y),(x,y)+F(u,v))$, where $(u,v)\in\Omega$. The terms in x and y translate the bundle in ray space. A translation in ray space is a shear as a function of z in the world, which does not change the cross section of the bundle at a fixed z. Therefore, without loss of generality we will consider only $(x,y)=(0,0)$ - the rays integrated by the central pixel of the output image.

We would like to know at what depth these rays meet. The cross section of this bundle of rays on the aperture plane $(z=0)$ is Ω. At the world plane $(z=1)$ the bundle has cross section Ω transformed by the matrix F. Linearly interpolating, at depth d the transform is:

$$F'_d = dF + (1-d)I \quad (4)$$

The kernel of this matrix tells us if the rays meet at a point or line. Equivalently, via the transform $\mu = \frac{d-1}{d}$, we can use the kernel of:

$$F''_\mu = F - \mu I \quad (5)$$

This kernel is described by the eigenvalues and eigenvectors of F. We now consider all cases and present a taxonomy of focus, which mirrors that of perspective.

Focused Cameras. Conventional anastigmatic focus corresponds to point perspective. With eigenvalues equal to μ, $F=\mu I$, and the focus is at $\frac{1}{1-\mu}$. As μ approaches negative infinity, the depth in focus approaches the aperture plane, at $\mu=0$ the focus is on the world plane, and at $\mu=1$ the focus is at infinity. $\mu>1$ corresponds to a pseudoscopic camera. In this case the focus is beyond infinity, with rays that converge behind the aperture plane. These four cases are illustrated in figure 5.i-iv.

Astigmatic Cameras. Linear astigmatic focus is characterized by two different depths of focus in two different directions (figure 5.v). This corresponds to cross slit cameras. The eigenvalues fix the depths of focus, and the eigenvectors fix the directions which become sharp. Equivalently, the eigenvectors are normal to the orientation of the bokeh at that depth. One depth of focus may be at infinity, which corresponds to the pushbroom camera (Figure 5.vi).

Partially Afocal Cameras. With two equal eigenvalues and a deficient eigenspace, rays converge in one direction only. This is illustrated at a finite depth in figure 5.vii, which corresponds to the pencil camera, and at infinity in figure 5.viii, which corresponds to the twisted orthographic camera. Note how the bokeh shears from a ellipse into a line.

Afocal Cameras. Complex eigenvalues give a bundle of rays which may neck down, as they do in figure 5.ix, but do not intersect, and hence do not focus at any point. This corresponds to the bilinear camera.

Relationships Between Cameras. Exactly the same relationships hold between the kinds of focus as the kinds of

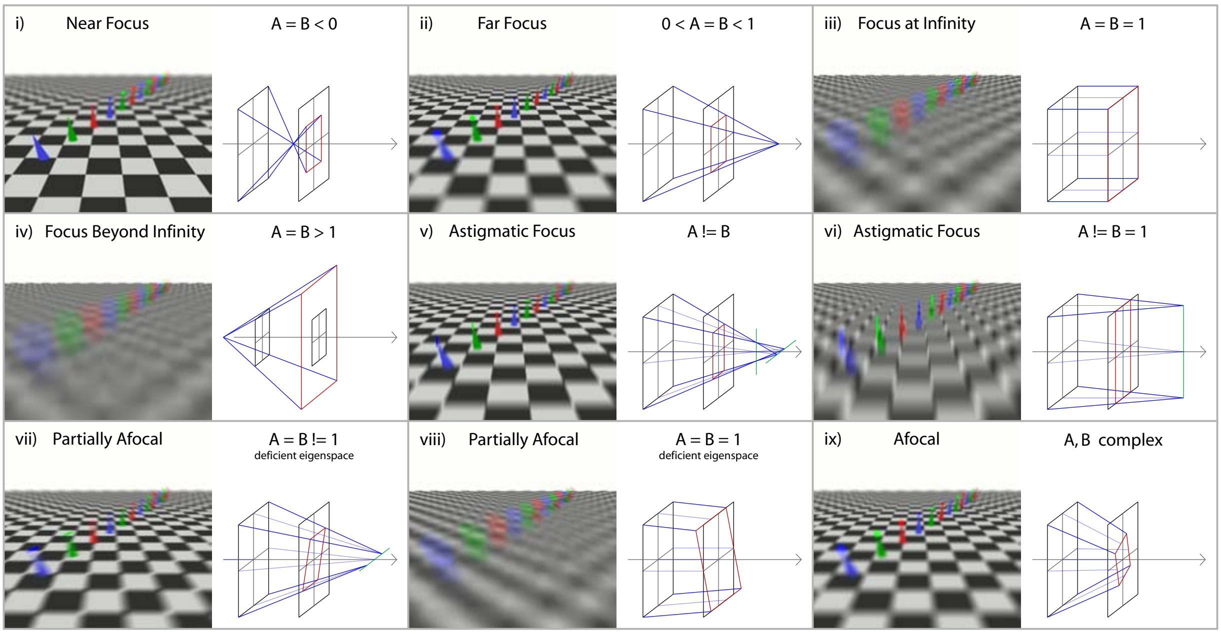

Figure 5: *Different conditions on the eigenvalues A, B of the matrix F give rise to different kinds of focus, in a manner that strictly parallels that of perspective (Figure 3). Each diagram shows a focused image of colored point lights resting atop cones on the left, and the rays integrated to produce the central pixel of that image on the right. The defocused point lights show how the bokeh changes with depth, and the blur on the checkerboard illustrates focus in the horizontal and vertical directions at each depth. In each line diagram, F is represented by the red parallelogram on the world plane.*

perspective. The general cases are afocal cameras and astigmatic cameras. Partially afocal cameras form the border of the two, and conventional anastigmatic focus is a special case of this. These relationships are illustrated on the right in figure 4.

5. General Linear Cameras with Finite Aperture

Most generally, a linear camera is a linear integral projection of ray space. Given a suitable choice of coordinate system, all such cameras can be expressed as the following integral over a light field L.

$$Im(x,y) = \int_{\Omega} L(Q(x,y,u,v))dudv \tag{6}$$

The 4x4 matrix Q can be factored as follows:

$$Q = \begin{pmatrix} A & PM \\ FA & M \end{pmatrix} = \begin{pmatrix} I & P \\ F & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M \end{pmatrix} \tag{7}$$

Two new matrices have appeared: A and M. We assume in this factorization that A and M are invertible. This rules out epipolar cameras, which explains why we omit them from the taxonomy. An epipolar camera images a one-dimensional subset of the world plane, and hence has a rank-one M.

The aperture matrix A warps the light field on the aperture plane. This space is integrated over, so A determines the shape of the bokeh. For example, $A = 2I$ is a large aperture and hence short depth of field, while $A = 0.1I$ produces a long depth of field. $A = 0$ produces a pinhole camera, and a rank-one A will produce a slit aperture. Note that due to the integration, not all A matrices have a unique effect. For example, pure rotations are equivalent to the identity.

In contrast, M is relatively uninteresting. It warps the light field on the (x,y) plane, which has no effect on the integration, so it just warps the output image. The perspective (P) and focus (F) matrices appear in this more general framework as shears between the (u,v) and (x,y) coordinates. Refocusing has been recognized to be a shear in ray space ([IMG00]), but its dual relationship with perspective has not been described explicitly before now.

This factorization demonstrates that F and P capture the interesting properties of a linear camera. All that remains is the aperture shape, determined by A, and the alignment of the output image, determined by M.

6. Real Time Implementation

To demonstrate this factorization, and to permit exploration of each branch of the taxonomies presented in sections 3 and 4, we have implemented a light field viewer for rendering discretely sampled light fields. For this purpose, equation 6 lends itself well to direct implementation in a fragment shader. The (x,y) coordinates are fixed by the fragment coordinates, and the (u,v) coordinates are iterated over in a loop inside the shader. Q is passed into the shader as a uniform 4x4 matrix, which warps (u,v,x,y) to determine a sample location. The light field, stored as a 3D texture, can then be trilinearly sample twice to perform the desired quadrilinear interpolation.

To implement a viewer that renders from a 3D model (as seen in figures 3 and 5), we use OpenGL with a vertex shader. For each (u,v) we render the scene, integrating the results in the accumulation buffer. Knowing Q and (u,v) we must transform each vertex v to the right screen space coordinates (x,y). We know that v must lie on the ray $(u',v',x',y') = Q(u,v,x,y)$, so:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = v_z \begin{pmatrix} x' \\ y' \end{pmatrix} + (1-v_z) \begin{pmatrix} u' \\ v' \end{pmatrix}$$

Expanding Q and solving for (x,y) yields:

$$\begin{aligned} K &= (v_z M + (1-v_z)PM)^{-1} \\ J &= K(v_z FA + (1-v_z)A) \\ \begin{pmatrix} x \\ y \end{pmatrix} &= K \begin{pmatrix} v_x \\ v_y \end{pmatrix} - J \begin{pmatrix} u \\ v \end{pmatrix} \end{aligned}$$

which can be implemented in a vertex program. The computations of K and J are dependent on v_z, and so must be done per vertex. This is not a linear map, so rational linear interpolation of attributes across polygons will not be correct. Our approximate solution is to subdivide large polygons finely enough that a linear approximation looks acceptable.

7. Future Work

This work reformulates and extends the concept of general linear cameras to include focus. Its appeal is its simplicity - that it is possible to understand general linear perspective and focus in terms of eigenspaces of 2x2 matrices. One possible future direction is to consider nonlinear integral projections that can be approximated locally as linear, with tangent P and F matrices at a given ray, analogously to the work of Yu and McMillan [YM05] [YM04a].

The main limitation of this framework is that once the two-planes are placed, all interesting events must be frontoparallel. A tilted focal plane requires a nonlinear projection of ray space, as does a cross slit camera with non frontoparallel slits. The most obvious mathematical extension is to add a third coordinate and use projective transforms, but this has not yet borne fruit. Some similarly simple extension of this framework to cover more general settings such as arbitrarily placed focal planes or slits would be interesting.

The challenge going forwards is to further generalize our notion of general linear cameras, without detriment to the concept's utility and elegance.

8. Acknowledgements

This work was partially funded by a Reed-Hodgson Stanford Graduate Fellowship.

References

[GH97] GUPTA R., HARTLEY R. I.: Linear pushbroom cameras. *IEEE Trans. Pattern Anal. Mach. Intell. 19*, 9 (1997), 963–975.

[Hal64] HALBACH K.: Matrix representation of gaussian optics. *American Journal of Physics 32* (feb 1964), 90–108.

[IMG00] ISAKSEN A., MCMILLAN L., GORTLER S. J.: Dynamically reparameterized light fields. In *Proc. SIGGRAPH '00* (2000), pp. 297–306.

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. SIGGRAPH '96* (1996), pp. 31–42.

[Paj02] PAJDLA T.: Stereo with oblique cameras. *Int. J. Comput. Vision 47*, 1-3 (2002), 161–170.

[RGL04] ROMAN A., GARG G., LEVOY M.: Interactive design of multi-perspective images for visualizing urban landscapes. In *Proc. VIS '04* (2004), pp. 537–544.

[VGT*05] VAISH V., GARG G., TALVALA E.-V., ANTUNEZ E., WILBURN B., HOROWITZ M., LEVOY M.: Synthetic aperture focusing using a shear-warp factorization of the viewing transform. In *Proc. Workshop on Advanced 3D Imaging for Safety and Security (A3DISS) 2005 (in conjunction with CVPR 2005)* (2005), p. 129.

[YM04a] YU J., MCMILLAN L.: A framework for multiperspective rendering. In *Eurographics Symposium on Rendering* (2004), pp. 61–68.

[YM04b] YU J., MCMILLAN L.: General linear cameras. In *ECCV (2)* (2004), pp. 14–27.

[YM05] YU J., MCMILLAN L.: Multiperspective projection and collineation. In *ICCV* (2005), pp. 580–587.

[ZFPW02] ZOMET A., FELDMAN D., PELEG S., WEINSHALL D.: Non-perspective imaging and rendering with the crossed-slits projection. In *Technical Report 2002-41* (2002), Leibnitz Center, Hebrew University of Jerusalem.