Washington University in St. Louis

[Washington University Open Scholarship](#)

Report Number: WUCS-01-42

2001-01-01

Relationship Between Two Generalized Images for Discrete and Differential Camera Motions

Robert Pless

The recent popularity of catadioptic and multi-camera imaging systems indicates a need to create formal models for general, non-perspective camera geometries. Development of algorithmic tools for interpreting images from a generalized camera model will lead to a better understanding of how to design camera systems for particular tasks. Here we define the corollary to epi-polar constraints for standard cameras - the relationship between two images of a scene taken by generalized cameras from viewpoints related by discrete or differential motions.

Follow this and additional works at: https://openscholarship.wustl.edu/cse_research

Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

[Department of Computer Science & Engineering](#) - Washington University in St. Louis
Campus Box 1045 - St. Louis, MO - 63130 - ph: (314) 935-6160.

Relationship Between Two Generalized
Images for Discrete and Differential Camera
Motions

Robert Pless

WUCS-01-42

December 2001

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
St. Louis MO 63130

Relationship between two generalized images for discrete and differential camera motions

Robert Pless

Abstract

The recent popularity of catadioptic and multi-camera imaging systems indicates a need to create formal models for general, non-perspective camera geometries. Development of algorithmic tools for interpreting images from a generalized camera model will lead to a better understanding of how to design camera systems for particular tasks. Here we define the corollary to epi-polar constraints for standard cameras — the relationship between two images of a scene taken by generalized cameras from viewpoints related by discrete or differential motions.

1   Introduction

The geometric relationships between several images taken of the same scene from different viewpoints has been the central subject of study in the Computer Vision community for many years [6]. The relationships have been formalized for orthographic, affine, and perspective projections. Recently new camera designs have appeared with different imaging geometries. For many purposes, these non-standard camera systems have significant advantages over pinhole or orthographic cameras.

The description and analysis of non-central projection camera systems ranges from very general imaging systems to those whose imaging geometry is highly constrained. The most general case allows an arbitrary mapping between image pixels and the part of the scene imaged by that pixel. This has been studied in the context of generating these images [11], and calibrating this imaging device [4]. More specialized camera geometries include linear push-broom cameras [5] and stereo panoramic cameras [9]. The equations of motion have been studied for a number of specific non-standard camera systems, including estimating optic flow from multi-camera systems [1, 2], and catadioptic systems [3].

Since new camera designs are constantly being proposed, it is useful to define a general framework for the analysis of non-central projection cameras which encompasses all new camera designs. The introduction and analysis of the
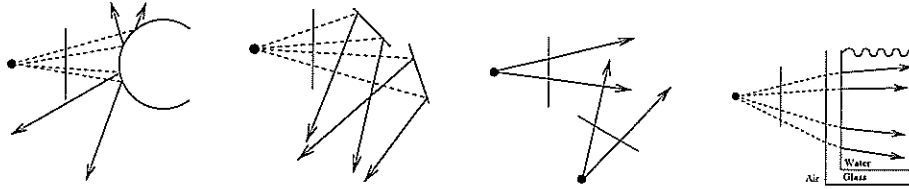
1

Figure 1: Examples of camera capturing systems within the framework of this paper. (left) A catadioptic system consisting of a camera looking at a curved mirror. (middle-left) An imaging system created with multiple flat mirrors. (middle-right) A rigidly mounted system of multiple synchronized cameras. (right) A scene being viewed through an optical interface which refracts the visible light. Note that the lines captured by these camera systems do not all pass through a single point (a non-central-projection).

oblique camera model [8] is one step in this direction, as it subsumes several of the existing non-standard camera models, including the linear push-broom and the stereo panoramic cameras. The oblique camera model is limited to camera geometries where a point in the world is imaged at most once by the camera — a system where none of the rays that are captured intersect.

In this work we remove all restrictions on the camera model, and provide an analysis of the geometry relating two views from the most generalized camera model, as considered by [11, 4]. This camera model encompasses all of the imaging situation depicted in Figure 1, multiple-camera systems, catadioptic imaging systems and cameras which view a scene through an optical interface. The only requirement is that there exists a one to one mapping between image pixels and the ray in space along which that pixel views. Specifically, there are no constraints on the projection model. In fact, the analysis does not even require a projection model; the geometric constraints rely only on the inverse projection model; a mapping from image pixels to rays in the scene. This mapping must be one to one, camera systems such as those depicted in Figure 2 are not considered. This mapping can be defined as a function of the image coordinates, or simply a list, explicitly defining, for each pixel, which ray in space it samples.

Calibrating this imaging system is significantly more difficult than calibrating a pinhole camera system. Also, many assumptions that are valid for normal camera geometries may not be valid for the arbitrary sensor geometries captured within this general framework. There may be discontinuities in the image capture process. Sampling issues which may be safely ignored for normal cameras may cause problems in the image analysis from general cameras. Standard feature detectors may fail because different parts of the image may have very different sampling properties. Classifying situations when sampling issues present problems needs to be considered on an application specific basis. This work is intended only to consider the geometric relationships, and to provide a general framework for the analysis of new camera designs.
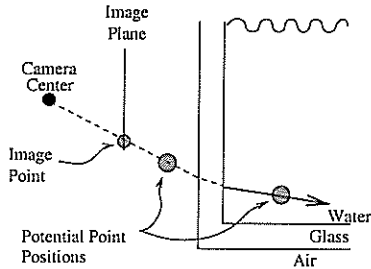
2

Figure 2: Environments without a one to one mapping between pixel coordinates and are the only imaging systems not included in this framework. For instance, from figure 1d, if an object can appear either before or after the optical interface, it is not known which line the pixel is sampling.

The following section describes some necessary background on the definition and calibration of the general imaging model. Then the relationships between two views of a static scene are illustrated, first for the case of the camera undergoing a discrete motion, in Section 3, and then for the case of differential motion, in Section 4. Section 5 gives two examples of real imaging systems considered in this generalized model; the discrete motion of a camera looking through an optic interface, and discrete and differential measurements from the bizarre eye of a stomatopod.

2   General Imaging Model

The general imaging model abstracts away from exactly what path light takes as it passes through the lenses and mirrors of an arbitrary imaging system. Instead, it identifies each image sensor reading with the region of space that affects that sensor. A reasonable model of this region of space is a cone emanating from some point. A complete definition of the imaging model has been defined in terms of "raxels" [4], (see Figure 3).

A raxel includes the following information about how a particular pixel samples the scene. This sampling is assumed to be centered around a ray starting at a point X,Y,Z, with a direction parameterized by (ϕ, θ). This pixel captures light from a cone around that ray, whose aspect ratio and orientation is given by (f_a, f_b, Υ). The light intensity captured may also be attenuated, these radiometric quantities may also differ for every pixel.

For the geometric analysis of multiple images, we simplify this calibration so that it only includes the definition of the ray that the pixel samples. This gives a simpler calibration problem which requires determining, for each pixel, the Plücker vectors of the sampled line. Since Plücker vectors are required for the mathematical analysis presented later, the following section gives a brief introduction.
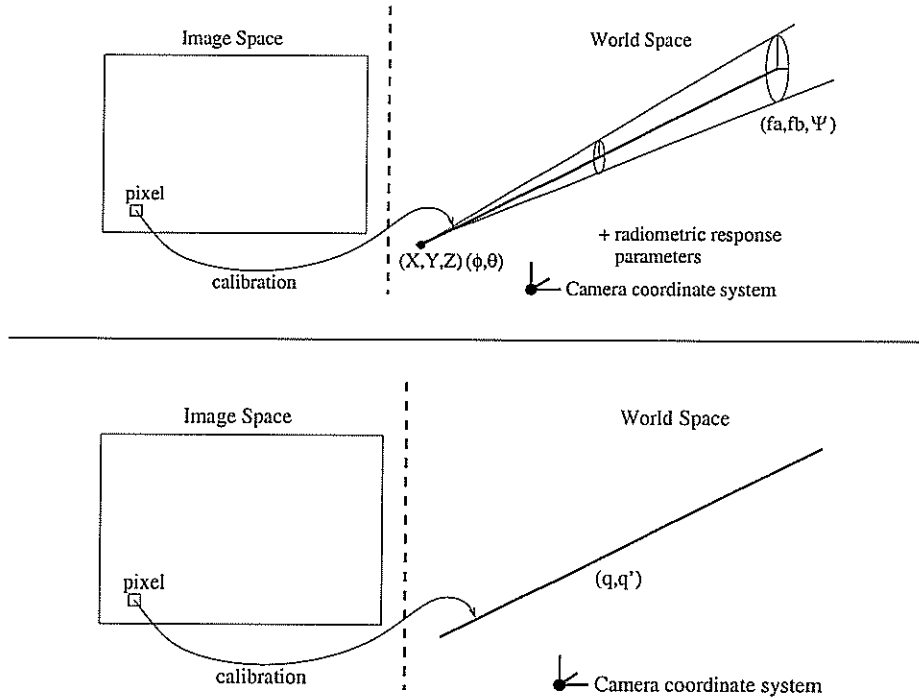
3

Figure 3: (Top) The generalized imaging system defined in [4] expresses, for each pixel, a description of how that pixel samples the light-field. This sampling is assumed to be centered around a ray starting at a point X,Y,Z, with a direction parameterized by (ϕ, θ), relative to a coordinate system attached to the camera. The pixel captures light from a cone around that ray, whose aspect ratio and orientation is given by (f_a, f_b, Υ). The light intensity captured may also be attenuated, these radiometric quantities may also differ for every pixel. (Bottom) The simplified imaging model parameterizes only the ray along which the scene is sampled, and does not consider radiometric properties. The ray is parameterized by its Plücker vectors q, q'.

2.1 Plücker Vectors

In order to describe the line in space that each pixel samples in this more general camera setting, we need a mechanism to describe arbitrary lines in space. There are many parameterizations of lines, but Plücker vectors [10] give a convenient mechanism for the types of transformations that are required. The Plücker vectors of a line are a pair of 3-vectors: q, q', named the direction vector and moment vector. q is a vector of any length in the direction of the line. Then, for any point P on the line $q' = q \times P$ There are two constraints that this pair of vectors must satisfy. First, $q \cdot q' = 0$, and second, the remaining five parameters are homogeneous, their overall scale does not affect which line they describe. It is often convenient to force the direction vector to be a unit vector, which defines a scale for the homogeneous parameters.

If q is a unit vector, the point $(q \times q')$ is the point on the line (defined by q, q') closest to the origin. The set of all points that lie on a line with these Plücker vectors is given by:

$$(q \times q') + \alpha q, \forall \alpha \in R \tag{1}$$

2.2 Plücker Calibration

The simplified calibration is an arbitrary mapping between pixel coordinates and the ray captured by that pixel. The calibration function is the mapping between each pixel captured by the camera and the line of sight captured by that pixel. This line is defined by its Plücker vectors. The calibration of this camera system defines the mapping between pixel coordinates and lines. The line captured by pixel (x, y) is $\langle q(x,y), q'(x,y) \rangle$. Most of the remainder of this paper discusses the constraint given by corresponding points in two images, or the optic flow at a point. In this case we drop the coordinates (x,y), and list the Plucker vectors simply as q, q'.

The analysis of differential camera motion is only possible if the imaging system smoothly samples the environment — that is if the ray sampled by one pixel is close to the ray sampled by neighboring pixels. In these cases, it is possible to define how the Plücker vectors of the sampled ray change with respect to a change of location in the image. That is to say, the following quantities are also defined (and computed as part of the calibration process).

$$\frac{\partial q(x,y)}{\partial x}, \frac{\partial q(x,y)}{\partial y}, \frac{\partial q'(x,y)}{\partial x}, \frac{\partial q'(x,y)}{\partial y}$$

These are four vector quantities, for intuition, in a standard normalized pinhole camera, $q'(x, y)$ is uniformly zero, $\frac{\partial q(x,y)}{\partial x} = \langle 1, 0, 0 \rangle$, and $\frac{\partial q(x,y)}{\partial y} = \langle 0, 1, 0 \rangle$. The calibration for the more generalized imaging system discussed in [4] (finding all of the parameters of the raxel) suffices to define the parameters of the Plücker Calibration. Additionally, the main contribution of this work is a theoretical model which encompasses many different specialized camera geometries. For any specific such geometry, the calibration process can be simplified. Given

5

this background, we can now consider the relationships between multiple images captured by these generalized cameras.

3   Discrete Motion

Suppose, in two generalized images, we have a correspondence between pixel (x_1, y_1) in the first image and pixel (x_2, y_2) in a second image. In our camera system, these points are projections of a 3d point which lies along lines in space described by Plücker vectors $\langle q_1, q_1' \rangle$, and $\langle q_2, q_2' \rangle$. The camera viewpoints are related by an arbitrary rigid transformation. There is a rotation R and a translation T which takes points in the coordinate system of the first camera and transforms them into the coordinate system of the second camera.

After this rigid transformation, the Plücker vectors of the first line in the second coordinate system become:

$$\langle Rq_1, Rq_1' + R(T \times q_1) \rangle \tag{2}$$

Since we have a pair of corresponding points, this line must intersect the line defined by the pixel coordinates of the corresponding point in the second camera. A pair of lines with Plücker vectors $\langle q_a, q_a' \rangle$, and $\langle q_b, q_b' \rangle$ intersect if and only if:

$$q_b \cdot q_a' + q_b' \cdot q_a = 0. \tag{3}$$

This allows us to write down the constraint given by the correspondence of a point between two images, combining Equations 2 and 3

$$q_2 \cdot (Rq_1' + R(T \times q_1)) + q_2' Rq_1 = 0.$$

This completely defines how two views of a point constrain the discrete motion of a generalized camera (using the convention that $[T]_x$ is the skew symmetric matrix such that $[T]_x v = T \times v$ for any vector v):

Generalized Epi-polar Constraint

$$q_2{}^T Rq_1' + q_2^T R[T]_x q_1 + q_2'{}^T Rq_1 = 0. \tag{4}$$

For standard perspective projection cameras, $q_1' = q_2' = 0$, and what remains: $q_2^T R[T]_x q_1 = 0$, is the classical epi-polar constraint defined by the Essential matrix.

Given the camera transformation R, T and corresponding points, it is possible to determine the 3D coordinates of the world point in view. Using Equation 1, and transforming the first line into the second coordinate system, solving

for the position of the point in space amounts to finding the intersection of the corresponding rays. This requires solving for the parameters α_1, α_2, which is the corollary of the depth in typical cameras:

$$R((q_1 \times q_1') + \alpha_1 q_1) + T = (q_2 \times q_2') + \alpha_2 q_2$$

Collecting terms leads to the following vector equation whose solution allows the reconstruction of scene point P in the coordinate system of the first camera.

Generalized Point Reconstruction

solve for α_1: $\alpha_1 R q_1 - \alpha_2 q_2 = (q_2 \times q_2') - R(q_1 \times q_1') - T,$

$$P = q_1 \times q_1' + \alpha q_1$$

---

4   Differential Motion

In the differential case, we will consider the image of a point P in space which is moving with a translation velocity $\vec{t}$, and an angular velocity, (relative to origin of the camera coordinate system) $\vec{\omega}$. The instantaneous velocity of the 3D point is:

$$\dot{P} = \vec{\omega} \times P + \vec{t}.$$

For a point in space to lie on a line with Plücker vectors $\langle q, q' \rangle$, the following must hold true:

$$P \times q - q' = \vec{0};$$

As the point moves, the Plücker vectors of the lines incident upon that point change. Together, the motion of the point and the change in the Plücker vectors of the line incident on that point must obey:

$$\frac{\partial}{dt}(P \times q - q') = \vec{0}, \text{or,}$$
$$\dot{P} \times q - P \times \dot{q} - \dot{q}' = 0$$

In terms of the parameters of motion, this gives:

$$(\vec{\omega} \times P + \vec{t}) \times q + P \times \dot{q} - \dot{q}' = \vec{0}.$$
$$(\vec{\omega} \times P + \vec{t}) \times q = \dot{q}' - P \times \dot{q}. \tag{5}$$

Which are constraints relating the camera motion and the line coordinates incident on a point in 3D for a particular rigid motion. On the image plane, the

image of this point is undergoing a motion characterized by its optic flow, (u, v) on the image plane. Combining this optic flow with the camera calibration, one can calculate how the coordinates of the Plücker vectors must be changing:

$$\dot{q} = \frac{\partial q}{\partial x}u + \frac{\partial q}{\partial y}v,$$

$$\dot{q}' = \frac{\partial q'}{\partial x}u + \frac{\partial q'}{\partial y}v, \tag{6}$$

so that we can consider $\dot{q}$ and $\dot{q}'$ to be image measurements. Then we can substitute Equation 1 into Equation 5 to get:

$$(\vec{\omega} \times ((q \times q') + \alpha q) + \vec{t}) \times q = \dot{q}' - ((q \times q') + \alpha q) \times \dot{q}.$$

or,

$$(\vec{\omega} \times (q \times q')) \times q + \alpha(\vec{\omega} \times q) \times q + \vec{t} \times q = \dot{q}' - (q \times q') \times \dot{q} + \alpha q \times \dot{q}. \tag{7}$$

The following identities hold when $|q| = 1$,

$$(\vec{\omega} \times (q \times q')) \times q = (\vec{\omega} \cdot q)(q \times q'), \text{ and,}$$
$$(\vec{\omega} \times q) \times q) = (\vec{\omega} \cdot q)q - \vec{\omega},$$
$$(q \times q') \times \dot{q} = -(q' \cdot \dot{q})q,$$

Simplifying Equation 7 and collecting terms gives:

$$\alpha((\vec{\omega} \cdot q)q - \vec{\omega} - q \times \dot{q}) + \vec{t} \times q = \dot{q}' + (q' \cdot \dot{q})q - (\vec{\omega} \cdot q)(q \times q') \tag{8}$$

Multiplying by "$\times q$", this becomes:

$$\alpha(\vec{\omega} \times q) + (\vec{t} \times q) \times q = \dot{q}' \times q + (\vec{\omega} \cdot q)q' + \alpha \dot{q}.$$

Collecting the terms that relate to the distance of the point along the sampled ray gives, and then dividing by α gives:

$$(\vec{\omega} \times q - \dot{q}) = \frac{-(\vec{t} \times q) \times q + \dot{q}' \times q + (\vec{\omega} \cdot q)q'}{\alpha}$$

which can be written more cleanly to define the optic flow for a generalized camera under differential motion:

$$\boxed{\begin{array}{c} \textbf{Generalized Optic Flow Equation} \\ \\ \dot{q} = \vec{\omega} \times q + \dfrac{(\vec{t} \times q) \times q - \dot{q}' \times q - (\vec{\omega} \cdot q)q'}{\alpha} \end{array}} \tag{9}$$

For standard perspective projection cameras, $q' = 0$, and $\dot{q}' = 0$, and this simplifies to the standard optic flow equation (for spherical cameras):

$$\dot{q} = (\vec{\omega} \times q) + \frac{(\vec{t} \times q) \times q}{\alpha}$$

To solve for the camera motion it is useful to find an expression relating the optic flow to the motion parameters that does not include the depth of the point. This differential form of the epi-polar constraint is:

$$\dot{q} \times ((\vec{t} \times q) \times q) = (\vec{\omega} \times q) \times ((\vec{t} \times q) \times q).$$

This same process can be applied to generalized cameras, giving:

Generalized Differential Epi-polar Constraint

$$(\dot{q} - \vec{\omega} \times q) \times ((\vec{t} \times q) \times q - \dot{q}' \times q - (\vec{\omega} \cdot q)q') = 0, \qquad (10)$$

which, like the formulation for planar and spherical cameras, is bilinear in the translation and the rotation and independent of the depth of the point.

5 Applications

In this section we illustrate the two view geometry for two different imaging systems that fall in the generalized camera model. The first involves a camera looking through an optic interface, the second is a simulation of the eye of a shrimp (stomatopod). Each of these imaging systems is a non-central projection.

5.1 Vision through optical interfaces

Many visual measurement processes, especially measurement tasks in various physical sciences, require a camera to view a scene of interest through an optical interface. Light refracts as it passes through an optical interface, so the standard pinhole model does not model the set of rays, in the scene of interest, that the camera is capturing.

The calibration of a view through an optic interface is highly application dependent. As an example application we considered a camera that is imaging a lava lamp (an toy consisting of a glass jar with colored wax and water which is heated from below. The heat causes globules of wax to rise — where they cool and subsequently sink). The intrinsic camera calibration was calculated from many images of a planar checkerboard pattern [12]. This checkerboard was then
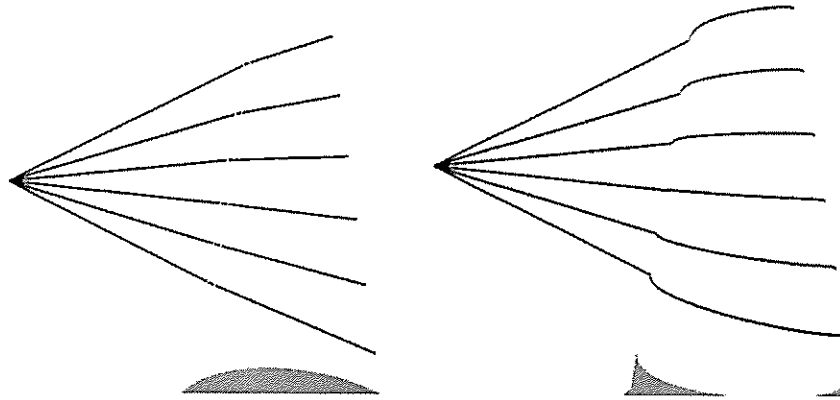
9

Figure 4: Left: The rays imaged by the camera at the left are refracted at the optic interface. Projecting these rays back through the optic interface to another camera position defines the epi-polar geometry for this camera pair.

augmented to include an arm which fit inside the lava lamp at a known relative position from the checkerboard. An image of the checkerboard with the (now) calibrated camera allows for the solution of the position of the augmenting arm in 3D. Eight thousand images of this form were taken. This was sufficient so that 17% of the pixels imaging the lava lamp had 2 or more samples of 3D points, sufficient to define the ray in the lava lamp along which they view. The calibration data was interpolated to extend to the remaining pixels which did not capture multiple 3D points in the calibration data set.

The calibration data was used to create a simulated views of the lava lamp. Figure 6 shows two views of the lava lamp, a point in the left view and the corresponding epi-polar curve in the right image.

5.2 Analysis of biological vision systems

Biological vision systems exhibit an astonishing diversity of design and function. Many mammals have eyes that are similar to standard video cameras, so pinhole camera model is a reasonable model for the imaging process. Insects, on the other the hand, tend to have more panoramic vision. For some insects this is relatively close to a central projection and the eye approximates the mathematical form of a spherical camera. For other systems, however, the view cannot be modeled by any central projection. One example of this is the astounding eye of the stomatopod 7.

The set of rays captured by the stomatopod eye was estimated from published goniometer measurements [7]. A simulation of the view from one eye is shown in Figure 8. The strip along the middle of the eye is known to have 16 visual pigments (as opposed to human eyes, which have 3). One theory on
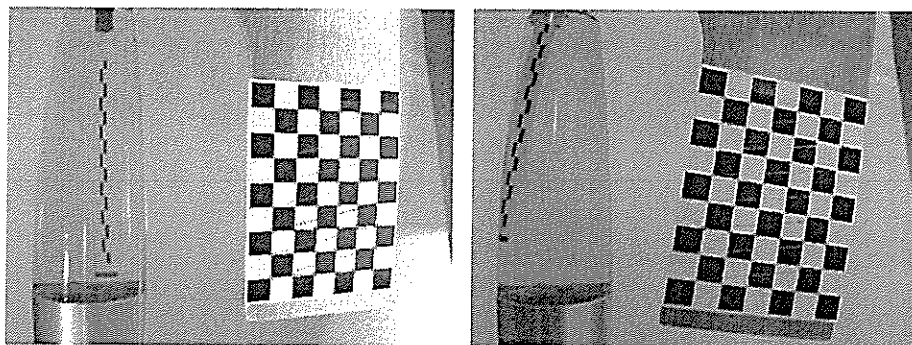
10

Figure 5: Two of the calibration images used to solve for the mapping from image pixels to rays *inside* the lava lamp, through the optic interface. The 3D position of the checkerboard is determined from its image in a calibrated camera, this defines the 3D position of the arm that is inside the lava lamp.

the formation of this eye is that the two hemispherical regions are optimized to estimate the eye rotation. As the eye rotates, the thin strip samples different parts of the visual field to paint or mosaic a high (color) resolution image of the environment. This is supported by reports of the natural motion of the shrimp eye which appears as a continual slow random scanning motion. Although it appears that data from all three regions may not be used to compute motion in this biological system, the equations given in Sections 3 and 4 would allow the analysis of either the two views or the flow field.

References

[1] P. Baker, R. Pless, C. Fermuller, and Y. Aloimonos. New eyes for shape and motion estimation. In *Biologically Motivated Computer Vision (BMCV2000)*, 2000.

[2] A Chaen, K Yamamzawa, N Yokoya, and H Takemura. Acquisition of three-dimensional information using omnidirectional stereo vision. In *Asian Conference on Computer Vision*, volume I, pages 288–295, 1998.

[3] Christopher Geyer and Kostas Daniilidis. Structure and motion from uncalibrated catadioptric views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[4] Micheal D Grossberg and Shree Nayar. A general imaging model and a method for finding its parameters. In *Proc. International Conference on Computer Vision*, volume II, pages 108–115, 2001.

[5] R Gupta and R Hartley. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):963–975, 1997.
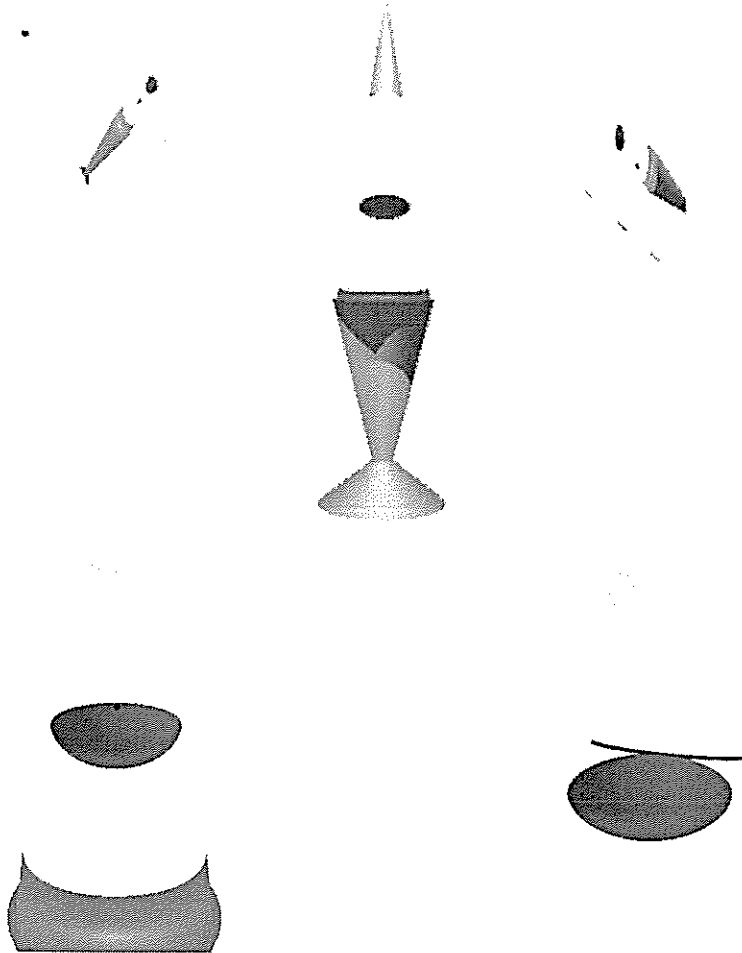
Figure 6: Two artificially generated views of the lava lamp, with the epi-polar geometry shown in the bottom images. Without an explicit projection model it is necessary to search through the rays in the second image to find those that (nearly) intersect with the ray captured by the specified point in the first image.
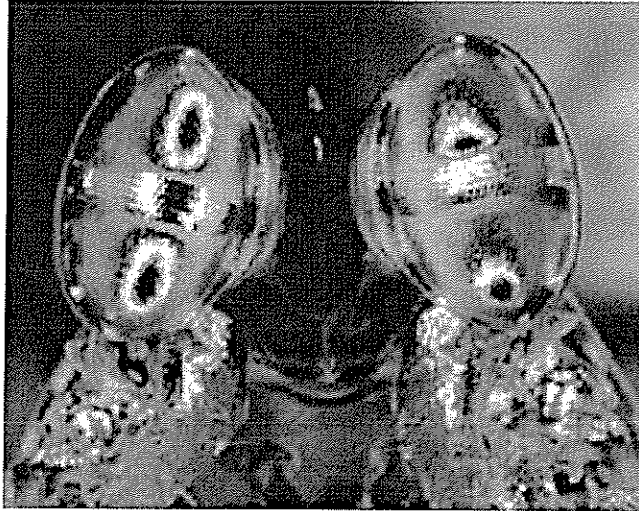
Figure 7: The compound eyes of stomatopods are on independently movable stalks. Each eye has three regions, two roughly hemispherical outer regions, separated by a roughly cylindrical strip between them which has up to 16 visual pigments. In this image, the reflection of the camera is visible as black surrounded by white, and is imaged three times by each eye.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 1999.

[7] N J Marshall and M F Land. Some optical features of the eyes of stomatopods: I. Eye shape, optical axes and resolution. *Journal of Comparitive Physiology, A*, 173:565–582, 1993.

[8] Tomas Pajdla. Characterization of epipolar geometries of non-classical cameras. Technical Report CTU-CMP-2001-05, Czech Technical University in Prague, December 2001.

[9] S Peleg, Y Pritch, and M Ben-Ezra. Cameras for stereo panoramic imaging. In *CVPR*, pages 208–214, 2000.

[10] J. Plücker. On a new geometry of space. *Philisophical Transactions of the Royal Society of London*, 155:725–791, 1865.

[11] P Rademacher and G Bishop. Multiple-center-of-projection images. In *SIGGRAPH*, 1998.

[12] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
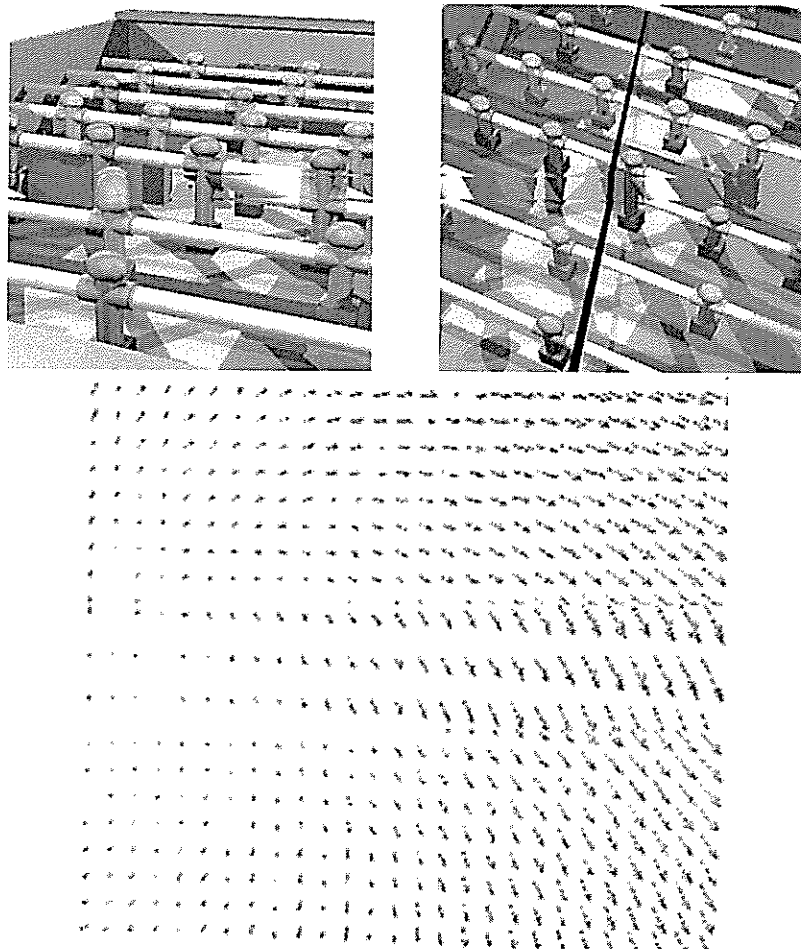
13

Figure 8: (top) Epi-polar geometry for two views from a simulated stomatopod eye, and (bottom) The optic flow field for a rigid motion while imaging a fronto-parallel plane.