Computing Matched-epipolar Projections

Richard Hartley and Rajiv Gupta
GE - Corporate Research and Development,
P.O. Box 8, Schenectady, NY, 12301.
Ph : (518)-387-7333
Fax : (518)-387-6845
email : hartley@crd.ge.com

Abstract

This paper gives a new method for image rectification, the process of resampling pairs of stereo images taken from widely differing viewpoints in order to produce a pair of "matched epipolar projections". These are projections in which the epipolar lines run parallel with the x-axis and consequently, disparities between the images are in the x-direction only. The method is based on an examination of the essential matrix of Longuet-Higgins which describes the epipolar geometry of the image pair. The approach taken is consistent with that recently advocated strongly by Faugeras ([1]) of avoiding camera calibration. The paper uses methods of projective geometry to define a matrix called the "epipolar transformation matrix" used to determine a pair of 2D projective transforms to be applied to the two images in order to match the epipolar lines. The advantages include the simplicity of the 2D projective transformation which allows very fast resampling as well as subsequent simplification in the identification of matched points and scene reconstruction.

1 Introduction

A recent paper [4] described an approach to stereo reconstruction that avoids the necessity for camera calibration. In that paper it is shown that the the 3-dimensional configuration of a set of points is determined up to a collineation of the 3-dimensional projective space $\mathcal{P}^3$by their configuration in two independent views from uncalibrated cameras. This calibration-free approach to the stereo reconstruction, or structure-from-motion problem was independently discovered and strongly advocated in [1]. The general method relies strongly on techniques of projective geometry, in which configurations of points may be subject to projective transformations in both 2-dimensional image space and 3-dimensional object space without changing the projective configuration of the points. In [4] it is shown that the *essential matrix*, Q, defined by Longuet-Higgins ([3]) is a basic tool in the analysis of two related images. The present paper develops further the method of applying projective geometric, calibration-free methods to the stereo problem.

The paper [4] starts from the assumption that point matches have already been determined between pairs of images, concentrating on the reconstruction of the 3D point set. In the present paper the problem of obtaining point matches between pairs of images is considered. In particular we consider the problem of matching images taken from very different view-points such that perspective distortion and different view-point make corresponding regions look very different. The approach taken is consistent with the projective-geometrical methods advocated in [1] and [4].

The method developed in this paper is to subject both the images to a 2-dimensional projective transformation so that the epipolar lines match up and run horizontally straight across each image. This ideal epipolar geometry is the one that will be produced by a pair of identical cameras placed side-by side with their principal axes parallel. Such a camera arrangement may be called a rectilinear stereo frame. For an arbitrary placement of cameras, however, the epipolar geometry will be more complex. In effect, transforming the two images by the appropriate projective transforms reduces the problem to that of a rectilinear stereo frame. Many stereo algorithms described in previous literature have assumed a rectilinear or near-rectilinear stereo frame.

After the 2D projective transformations have been applied to the two images, matching points in the two images will have the same y-coordinate, since the epipolar lines match and are parallel to the x-axis. It is shown that the two transformations may be chosen in such a way that matching points have approximately the same x-coordinate as well. In this way, the two images, if overlaid on top of each other will correspond as far as possible, and any disparities will be parallel to the x-axis. Since the application of arbitrary 2D projective transformations may distort the image substantially, a method is described for finding a pair of transforms which subject the images to minimal distortion.

The advantages of reducing to the case of a rectilinear stereo frame are two-fold. First, the search for matching points is vastly simplified by the simple epipolar structure and by the near-correspondence of the two images. Second, a correlation-based match-point search can succeed, because local neighbourhoods around matching pixels will appear similar and hence will have high correlation.

The method of determining the 2D projective transformations to apply to the two images is new, making use of the essential matrix Q, and a new matrix defined here, which may be called the epipolar transformation matrix, denoted by M for the pair of images. If the method described in [4] is used for the 3D reconstruction of the object, then it is not necessary to take account of the effect of the 2D transformations. In fact, once the two images have been transformed, the original images may be thrown away and the transforms forgotten, since in the first instance, the object will be reconstructed up to a 3D collineation only. As in [4] it is necessary to use ground-control points, or some other constraints to reconstruct the absolute configuration of the scene (up to a similarity transform). Because we are effectively dealing with a rectilinear stereo frame, the mathematics of the reconstruction of the 3D points is extremely simple.

1.1 Preliminaries

The symbol $\mathbf{u}$ represents a column vector. We will use the letters u, v and w for homogeneous coordinates in image-space. In particular, the symbol $\mathbf{u}$ represents the column vector $(u, v, w)^\top$. Object space points will also be represented by homogeneous coordinates x, y, z and t, or more often x, y, z and 1. The symbol $\mathbf{x}$ will represent a point in three-dimensional projective space represented in homogeneous coordinates.

Since all vectors are represented in homogeneous coordinates, their values may be multiplied by any arbitrary non-zero factor. The notation $\approx$ is used to indicate equality of vectors or matrices up to multiplication by a scale factor.

If A is a square matrix then its matrix of cofactors is denoted by A^*. The following identities are well known : $A^*A = AA^* = \det(A)I$ where I is the identity matrix. In particular, if A is an invertible matrix, then $A^* \approx (A^\top)^{-1}$.

Given a vector, $\mathbf{t} = (t_x, t_y, t_z)^\top$ it is convenient to introduce the skew-symmetric matrix

$$[\mathbf{t}]_\times = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \quad (1)$$

For any non-zero vector $\mathbf{t}$, matrix $[\mathbf{t}]_\times$ has rank 2. Furthermore, the null-space of $[\mathbf{t}]_\times$ is generated by the vector $\mathbf{t}$. This means that $\mathbf{t}^\top[\mathbf{t}]_\times = [\mathbf{t}]_\times\mathbf{t} = 0$ and that any other vector annihilated by $[\mathbf{t}]_\times$ is a scalar multiple of $\mathbf{t}$.

The matrix $[\mathbf{t}]_\times$ is closely related to the cross-product of vectors in that for any vectors $\mathbf{s}$ and $\mathbf{t}$, we have $\mathbf{s}^\top[\mathbf{t}]_\times = \mathbf{s} \times \mathbf{t}$ and $[\mathbf{t}]_\times\mathbf{s} = \mathbf{t} \times \mathbf{s}$. A useful property of cross products may be expressed in terms of the matrix $[\mathbf{t}]_\times$.

Proposition 1.1. *For any 3×3 matrix M and vector $\mathbf{t}$*

$$M^*[\mathbf{t}]_\times = [M\mathbf{t}]_\times M \quad (2)$$

Camera Model. The general model of a perspective camera that will be used here is that represented by an arbitrary 3×4 matrix, P, known as the *camera matrix*. The camera matrix transforms points in 3-dimensional projective space to points in 2-dimensional projective space according to the equation $\mathbf{u} = P\mathbf{x}$. The camera matrix P is defined up to a scale factor only, and hence has 11 independent entries. This model allows for the modeling of several parameters, in particular : the location and orientation of the camera; the principal point offsets in the image space; and unequal scale factors in two orthogonal directions not necessarily parallel to the axes in image space.

If the camera is not placed at infinity, then the left-hand 3×3 submatrix of P is non-singular. Then P can be written as $P = (M \mid -M\mathbf{t})$ where $\mathbf{t}$ is a vector representing the location of the camera.

2 Epipolar Geometry

Suppose that we have two images of a common scene and let $\mathbf{u}$ be a point in the first image. It is well known that the matching point $\mathbf{u}'$ in the second image must lie on a specific line called the *epipolar line* corresponding to $\mathbf{u}$. The epipolar lines in the second image corresponding to all points $\mathbf{u}$ in the first image all meet in a point $\mathbf{p}'$, called the *epipole*. The epipole $\mathbf{p}'$ is the point where the centre of projection of the first camera would be visible in the second image. Similarly, there is an epipole $\mathbf{p}$ in the first image defined by reversing the roles of the two images in the above discussion.

Points in an image will be represented by homogeneous coordinates. For instance $\mathbf{u} = (u, v, w)^\top$ represents a point in the first image. Similarly, lines in an image are represented by a homogeneous vector. A point $\mathbf{u}$ lies on a line $\mathbf{l}$ if and only if $\mathbf{l}^\top \mathbf{u} = 0$. It was shown in [3, 2, 5] that there exists a 3×3 matrix Q called the *essential matrix* which maps points in one image to the corresponding epipolar line in a second image according to the mapping $\mathbf{u} \mapsto Q\mathbf{u}$. If $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ are a set of matching points, then the fact that $\mathbf{u}'_i$ lies on the epipolar line $Q\mathbf{u}_i$ means that

$$\mathbf{u}'^\top_i Q \mathbf{u}_i = 0 \ . \qquad (3)$$

Given at least 8 point matches, it is possible to determine the matrix Q by solving a set of linear equations of the form (3).

The properties of the essential matrix will be given below in a number of theorems. For proofs, the reader is referred to [5] or [4]. The first proposition states some simple properties of the essential matrix.

Proposition 2.2. *Suppose that Q is the essential matrix corresponding to an ordered pair of images (J, J').*

1. *Matrix $Q^\top$ is the essential matrix corresponding to the ordered pair of images (J', J).*
2. *The epipole $\mathbf{p}$ is the unique point such that $Q\mathbf{p} = 0$*
3. *Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be points in J, neither of which is equal to the epipole $\mathbf{p}$. Points $\mathbf{u}_1$ and $\mathbf{u}_2$ are on the same epipolar line if and only if $Q\mathbf{u}_1 \approx Q\mathbf{u}_2$.*
4. *Q factors as a product $Q = M^*[\mathbf{p}]_\times$ for some non-singular matrix M.*

In using the phrase "unique point such that $Q\mathbf{p} = 0$", the solution $\mathbf{p} = (0, 0, 0)^\top$ is excluded as not being a valid set of homogeneous coordinates. Further, since we are dealing in homogeneous coordinates, non-zero scale factors are insignificant. Note that in part 4 of Proposition 2.2, we could equally well write $Q = M[\mathbf{p}]_\times$ for some non-singular matrix, M. However, the form given in the proposition is preferred. As shown by Proposition 2.2, the matrix Q determines the epipoles in both images. Furthermore, Q provides the map between points in one image and epipolar lines in the other image. Proposition 2.2 shows that this map induces a correspondence between the epipolar lines in one image and the epipolar lines in the other image. Thus, the complete geometry and correspondence of epipolar lines is encapsulated in the essential matrix.

A formula for the essential matrix may also be given directly in terms of the two camera transformations ([5]).

Proposition 2.3. *The essential matrix corresponding to a pair of cameras with transform matrices $P = (K \mid -K\mathbf{t})$ and $P' = (K' \mid -K'\mathbf{t}')$ is given by the formula*

$$Q \approx (K'K^{-1})^*[K(\mathbf{t}' - \mathbf{t})]_\times \qquad (4)$$

If two transform matrices P and P' satisfy (4) for a given essential matrix Q, we say that Q is realized by the pair (P, P'). If Q is written as $Q = M^*[\mathbf{p}]_\times$ then it is easily verified that one realization of Q is given by the camera pair
$\{(I \mid 0), (M \mid -M\mathbf{p})\}$.

The factorization of Q into a product of non-singular and skew-symmetric matrices is not unique, as is shown by the following proposition.

Proposition 2.4. *Let the 3×3 matrix Q factor in two different ways as $Q \approx M_1 S_1 \approx M_2 S_2$ where each S_i is a non-zero skew-symmetric matrix and each M_i is non-singular. Then $S_2 \approx S_1$. Furthermore, if we write $S_i = [\mathbf{p}]_\times$ then $M_2 \approx M_1 + \mathbf{a}\mathbf{p}^\top$ for some vector $\mathbf{a}$.*

As a consequence of Proposition 2.4, the realization of an essential matrix is not unique. The following theorem shows how different realizations of the same essential matrix are related.

Theorem 2.5. *Let $\{P_1, P'_1\}$ and $\{P_2, P'_2\}$ be two pairs of camera transforms. Then $\{P_1, P'_1\}$ and $\{P_2, P'_2\}$ correspond to the same essential matrix Q if and only if there exists a 4×4 non-singular matrix H such that $P_1 H \approx P_2$ and $P'_1 H \approx P'_2$.*

The following basic theorem showing that the knowledge of the epipolar correspondence determines the reconstructed scene geometry up to a 3D collineation follows easily from Theorem 2.5.

Theorem 2.6. (Faugeras [1], Hartley et al. [4]) *If the essential matrix Q for a pair of images is known and if $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ is a set of matched points, then the locations $\mathbf{x}_i$ of the corresponding 3D points are determined up to a collineation of $\mathcal{P}^3$.*

3 The epipolar transform matrix

Given a factorization of an essential matrix, $Q = M[\mathbf{p}]_\times$, we show that the matrix M is of interest in its own right.

The following proposition lists some of the properties of the matrix M.

Proposition 3.7. *Let Q be an essential matrix and $\mathbf{p}$ and $\mathbf{p}'$ the two epipoles. If Q factors as a product $Q = M^*[\mathbf{p}]_\times$ then*

1. $M\mathbf{p} = \mathbf{p}'$.

2. *For any epipolar line* $\mathbf{l}$ *in the first image,* $M^*\mathbf{l}$ *is the corresponding epipolar line in the second image.*

Proof.

1. Since $Q \approx M^*[\mathbf{p}]_\times$, it follows that $\mathbf{p}^\top M^\top Q \approx \mathbf{p}^\top M^\top M^*[\mathbf{p}]_\times = \mathbf{p}^\top[\mathbf{p}]_\times = 0$. Therefore, $Q^\top(M\mathbf{p}) = 0$ and it follows from Proposition (2.2) that $M\mathbf{p} \approx \mathbf{p}'$.

2. Let $\mathbf{u}$ be a point in the first image and $\mathbf{l}$ be the corresponding epipolar line in the second image. Thus, $\mathbf{l}' = Q\mathbf{u}$. The epipolar line through $\mathbf{u}$ in the first image is $\mathbf{l} = \mathbf{p}\times\mathbf{u} = [\mathbf{p}]_\times\mathbf{u}$. Consequently, $M^*\mathbf{l} = M^*[\mathbf{p}]_\times\mathbf{u} = Q\mathbf{u} = \mathbf{l}'$ as required.

□

Thus, M maps epipoles to epipoles and M^* maps epipolar lines to epipolar lines. This is true whichever matrix M is chosen as long as $Q = M^*[\mathbf{p}]_\times$.

3.1 M as a point map

As was seen in Proposition 3.7 if M is interpreted as a point-to-point map between images, then it maps one epipole to the other. In this section, other properties of M as a point map will be investigated and it will be shown that there is a natural one-to-one correspondence between factorizations of the essential matrix and planes in $\mathcal{P}^3$. This correspondence is expressed in the following theorem.

Theorem 3.8. *Let Q be the essential matrix corresponding to a pair of images represented by camera transforms P and P'. For each plane π in $\mathcal{P}^3$ not passing through the centre of projection of either camera there exists a factorization $Q = M^*[\mathbf{p}]_\times$ such that if $\mathbf{u} \leftrightarrow \mathbf{u}'$ is a pair of matched points corresponding to a point $\mathbf{x}$ in the plane π, then $\mathbf{u}' = M\mathbf{u}$. Conversely, for each factorization $Q = M^*[\mathbf{p}]_\times$ there exists a plane π with this property.*

Proof. By an appropriate choice of projective coordinates, it may be assumed that the plane π is the plane at infinity. Any point $\mathbf{x}$ in π is of the form $(x, y, z, 0)^\top$. The camera matrices may be written in the form $(K \mid L)$ and $(K' \mid L')$, and since the camera centres do not lie on the plane at infinity, the matrices K and K' are non-singular. According to (4) the essential matrix is $Q = (K'K^{-1})^*[\mathbf{p}]_\times$ and so we may choose $M = K'K^{-1}$. Now, point $\mathbf{u} = K(x, y, z)^\top$ and $\mathbf{u}' = K'(x, y, z)^\top$. It follows that $\mathbf{u}' = K'K^{-1}\mathbf{u} = M\mathbf{u}$ as required.

Conversely, suppose a factorization $Q = M^*[\mathbf{p}]_\times$ is given. One realization of Q is given by the camera pair $\{P_1, P_1'\} = \{(I \mid 0), (M \mid -M\mathbf{p})\}$. Let the plane π_1 be the plane at infinity, and suppose $\mathbf{x} = (x, y, z, 0)^\top$. Then $\mathbf{u} = P_1\mathbf{x} = (x, y, z)^\top$ and $\mathbf{u}' = P_1'\mathbf{x} = M(x, y, z)^\top$, and so $\mathbf{u}' = M\mathbf{u}$. Now, consider the camera pair $\{P, P'\}$. According to Theorem 2.5, there exists a non-singular 4×4 matrix, H such that $P = P_1H$ and $P' = P_1'H$. Then define π to be the plane $H^{-1}\pi_1$ and consider a point $H^{-1}\mathbf{x}$ in this plane Then, $PH^{-1}\mathbf{x} = P_1HH^{-1}\mathbf{x} = P_1\mathbf{x} = \mathbf{u}$, and similarly, $P'H^{-1}\mathbf{x} = \mathbf{u}' = M\mathbf{u}$. In other words, the plane π satisfies the requirements of the theorem. □

3.2 Computation of M.

We have just shown that for point sets that lie close to a plane π and are viewed from two cameras, there is a matrix M such that M^* maps an epipolar line in one image to the corresponding epipolar line in the second image, and such that M maps points $\mathbf{u}$ in the first image to a point close to the matched point $\mathbf{u}'$ in the second image. For matched points $\mathbf{u} \leftrightarrow \mathbf{u}'$ corresponding to a point $\mathbf{x}$ on the plane π the map $\mathbf{u} \mapsto M\mathbf{u} = \mathbf{u}'$ will be exact, whereas for points near to π the match will be approximate.

For instance, in the standard stereo-matching problem where the task is to find a pixel $\mathbf{u}'$ in a second image to match the pixel $\mathbf{u}$ in the first image, a good approximation to $\mathbf{u}'$ may be given by $M\mathbf{u}$. The best match, $\mathbf{u}'$ may be found by an epipolar search centred at the point $M\mathbf{u}$.

The goal of this section is, given a set of matched points $\mathbf{u}_i \leftrightarrow \mathbf{u}_i'$, to find the epipolar transformation matrix M which most nearly approximates the matched point correspondence. Specifically, given an essential matrix Q computed from the correspondences $\mathbf{u}_i \leftrightarrow \mathbf{u}_i'$, our goal is to find a projective transformation given by a matrix M, such that $M^\top Q$ is skew-symmetric and such that $\sum ||\mathbf{u}_i' - M\mathbf{u}||^2$ is minimized.

This may be considered as a constrained minimization problem. First, we consider the constraints. The condition that $M^\top Q$ should be skew-symmetric leads to a set of 6 equations in the 9 entries of M.

$$\begin{aligned}
m_{11}q_{11} + m_{21}q_{21} + m_{31}q_{31} &= 0 \\
m_{12}q_{12} + m_{22}q_{22} + m_{32}q_{32} &= 0 \\
m_{13}q_{13} + m_{23}q_{23} + m_{33}q_{33} &= 0 \\
m_{11}q_{12} + m_{21}q_{22} + m_{31}q_{32} &= -(m_{12}q_{11} + m_{22}q_{21} + m_{32}q_{31}) \\
m_{11}q_{13} + m_{21}q_{23} + m_{31}q_{33} &= -(m_{13}q_{11} + m_{23}q_{21} + m_{33}q_{31}) \\
m_{12}q_{13} + m_{22}q_{23} + m_{32}q_{33} &= -(m_{13}q_{12} + m_{23}q_{22} + m_{33}q_{32})
\end{aligned} \tag{5}$$

The entries q_{ij} of the matrix Q are known, so this gives a set of 6 known constraints on the entries of M. However, because of the fact that Q is singular of rank 2, there is one redundant restraint as will be shown next.

Next, consider the minimization of the goal function. We have a set of correspondences, $\{\mathbf{u}_i\} \leftrightarrow \{\mathbf{u}'_i\}$ and the task is to find the M that best approximates the correspondence. We can cast this problem in the form of a set of linear equations in the entries of M as follows. Let M_i be the i-th row of M. Then the basic equation $\mathbf{u}' = M\mathbf{u}$ can be written as

$$\begin{aligned} m_{11}u_i + m_{12}v_i + m_{13} &= w'_i u'_i \\ m_{21}u_i + m_{22}v_i + m_{23} &= w'_i v'_i \\ m_{31}u_i + m_{32}v_i + m_{33} &= w'_i \end{aligned}$$

We can substitute the third equation into the first two to get two equations linear in the m_{ij}

$$\begin{aligned} m_{11}u_i + m_{12}v_i + m_{13} &= m_{31}u_i u'_i + m_{32}v_i u'_i + m_{33}u'_i \\ m_{21}u_i + m_{22}v_i + m_{23} &= m_{31}u_i v'_i + m_{32}v_i v'_i + m_{33}v'_i \end{aligned} . \quad (6)$$

The method used here is essentially that described by Sutherland [6]. It does not minimize exactly the squared error $\sum ||M\mathbf{u}_i - \mathbf{u}'_i||^2$, but rather a sum weighted by w'_i. However, we accept this limitation in order to use a fast linear technique. This has not caused any problems in practice.

Since the matrix, M is determined only up to a scale factor, we seek an appropriately normalized solution. In particular, we seek a minimize the error in (6) subject to the constraints (5) and the condition $\sum m_{ij}^2 = 1$.

3.3 Solution of the constrained minimization problem.

Writing the constraint equations as $B\mathbf{x} = 0$ and the equations (6) as $A\mathbf{x} = 0$, our task is to find the solution $\mathbf{x}$ that fulfills the constraints exactly and most nearly satisfies the conditions $A\mathbf{x} = 0$. More specifically, our task is to minimize $||A\mathbf{x}||$ subject to $||\mathbf{x}|| = 1$ and $B\mathbf{x} = 0$. One method of solving this is to proceed as follows.

Extend B to a square matrix B' by the addition of 3 rows of zeros. Let the Singular Value Decomposition of B' be $B' = UDV^\top$, where V is a 9×9 orthogonal matrix and D is a diagonal matrix with 5 non-zero singular values, which may be arranged to appear in the top left-hand corner. Writing $\mathbf{x}' = V^\top\mathbf{x}$, and $\mathbf{x} = V\mathbf{x}'$, we see that $||\mathbf{x}|| = ||\mathbf{x}'||$. The problem now becomes, minimize $||AV\mathbf{x}'||$ subject to $UD\mathbf{x}' = 0$ and $||\mathbf{x}'|| = 1$. The condition that $UD\mathbf{x}' = 0$ means $D\mathbf{x}' = 0$, and hence the first five entries of $\mathbf{x}'$ are zero, since D is diagonal with its first five entries nonzero. Therefore, let A'' be the matrix formed from AV by dropping the first five columns. We now solve the problem : minimize $||A''\mathbf{x}''||$ subject to $||\mathbf{x}''|| = 1$. This is a straightforward unconstrained minimization problem and the solution $\mathbf{x}''$ is the singular vector corresponding to the smallest singular value of A''. Once $\mathbf{x}''$ is found, vector $\mathbf{x}'$ is obtained from it by appending 5 zeros. Finally, $\mathbf{x}$ is found according to the equation $\mathbf{x} = V\mathbf{x}'$.

4 Matched Epipolar Projections.

Consider the situation in which we have two images J_0 and J'_0 of the same scene taken from different unknown viewpoints. Suppose that the essential matrix, Q_0 and an epipolar transformation matrix M_0 have been computed based on a set of image-to-image correspondences between the two images. In general, the epipolar lines in the two images will run in quite different directions, and epipolar lines through different points will not be parallel (since they meet at the epipole). The goal in this next section is to define two perspective transformations, F and F', to be applied to the two images so that if the image J_0 is transformed according to the perspective transformation F, and J'_0 is transformed according to the perspective transformation F', then the resulting images J_1 and J'_1 correspond to each other in a particularly simple manner. In particular, the epipolar lines in the new images will be horizontal and parallel. Further, the resulting image-to-image transformation M_1 will be the identity mapping.

Let the epipole in the first image be $\mathbf{p}_0$. The epipolar lines all pass through $\mathbf{p}_0$ and hence are not parallel. Our goal is to transform the image so that the epipole is moved to the point $(1, 0, 0)^\top$ in homogeneous coordinates. This point is the point at infinity in the direction along the x axis. If this is the new epipole, then the epipolar lines will all be parallel with the x axis. Therefore, let F be a perspective transformation that sends the point $\mathbf{p}$ to $(1, 0, 0)^\top$. This is not by itself sufficient information to determine F uniquely, and F will be more exactly determined later.

Proposition 4.9. *Suppose there exist two images J_0 and J'_0 with corresponding essential matrix $Q = M^*[\mathbf{p}]_\times$. Let $\mathbf{p}$ and $\mathbf{p}'$ be the two epipoles. Let F be a homogeneous transformation such that $F\mathbf{p} = (1, 0, 0)^\top$ and let $F' = FM^{-1}$. Then, $F'\mathbf{p}' = (1, 0, 0)^\top$.*

Proof. $F'\mathbf{p}' = F.M^{-1}\mathbf{p}' = F\mathbf{p} = (0, 0, 1)^\top$. □

We now use the two projective transformations F and F' to resample the two images J_0 and J'_0 to give two new images J_1 and J'_1. By this is meant that J_1 is related to J_0 by the property that any point $\mathbf{x}$ in 3-space that is imaged at point $\mathbf{u}$ in image J_0 will be imaged at point $F\mathbf{u}$ in image $J_1 = F(J_0)$.

Proposition 4.10. *The essential matrix for the pair of resampled images J_1 and J_1' obtained by resampling according to the maps F and F' defined in the previous proposition is given by*

$$Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

The corresponding epipolar transformation matrix M_1 is the identity matrix.

Proof. Let $\{\mathbf{u}_i\} \leftrightarrow \{\mathbf{u}_i'\}$ be a set of image-to-image correspondences between the images J_0 and J_0' sufficient in number to determine the matrix Q_0 uniquely. Thus, $\mathbf{u}_i'^{\top} Q_0 \mathbf{u}_i = 0$. However, in the images J_1 and J_1', the points $\{F\mathbf{u}_i\}$ correspond with points $\{F'\mathbf{u}_i'\}$. It follows that

$$\mathbf{u}_i'^{\top} F'^{\top} Q_1 F \mathbf{u}_i = 0$$

and hence $Q_0 \approx F'^{\top} Q_1 F$ or

$$Q_1 = F'^{*} Q_0 F^{-1} \ . \tag{7}$$

However, since $F' = FM^{-1}$, we may write $F'^{*} = F^{*} M^{\top}$, and substituting in (7) gives

$$\begin{aligned} Q_1 &= F^{*} M^{\top} Q_0 F^{-1} \\ &= F^{*}[\mathbf{p}]_{\times} F^{-1} && \text{by Proposition 3.7} \\ &= F^{*} F^{\top} [F\mathbf{p}]_{\times} && \text{by (2)} \\ &= [(1,0,0)^{\top}]_{\times} \end{aligned}$$

which is the required matrix.

To prove the second statement, let $\mathbf{x}$ be a point in space on the plane determined by the image-to-image projective transform M in the sense that if $\mathbf{u}$ and $\mathbf{u}'$ are the images of point $\mathbf{x}$ in the images J_0 and J_0', then $\mathbf{u}' = M\mathbf{u}$ (see Proposition 3.8). Then point $\mathbf{x}$ will be seen in the images J_1 and J_1' at points $F\mathbf{u}$ and $F'\mathbf{u}'$. However, $F'\mathbf{u}' = F'M\mathbf{u} = F\mathbf{u}$. So, $F\mathbf{u}$ in image J_1 is mapped to $F'\mathbf{u}'$ in image J_1' by the identity transformation. □

4.1 Determination of the resampling transformation.

The transformation F was described by the condition that it takes the epipole $\mathbf{p}$ to the point at infinity on the x axis. This leaves many degrees of freedom open for F, and if an inappropriate F is chosen, severe projective distortion of the image can take place. In order that the resampled image should look somewhat like one of the original images, we may put closer restrictions on the choice of F.

One condition that leads to quite good results is to insist that the transformation F should act as far as possible as a rigid transformation in the neighbourhood of a given selected point $\mathbf{u}_0$ of the first image. By this is meant that the neighbourhood of $\mathbf{u}_0$ may undergo rotation and translation only, and hence will look the same in the original and resampled image. An appropriate choice of point $\mathbf{u}_0$ may be the centre of the image. For instance, this would be a good choice in an context of aerial photography if the first image is known not to be excessively oblique.

A projective transformation may be determined by specifying the destination of four points. Suppose that the epipole is already on the x axis at location $(1,0,f)^{\top}$ and that we desire the projective transformation to approximate the identity map in the local neighbourhood of the origin $(0,0,1)^{\top}$. The desired map may be found by specifying the destinations of four points

$$\begin{array}{lcl} (1,0,f)^{\top} & \rightarrow & (1,0,0)^{\top} \\ (0,0,1)^{\top} & \rightarrow & (0,0,1)^{\top} \\ (\delta,\delta,1)^{\top} & \rightarrow & (\delta,\delta,1)^{\top} \\ (\delta,-\delta,1)^{\top} & \rightarrow & (\delta,-\delta,1)^{\top} \end{array} \tag{8}$$

and then letting $\delta \rightarrow 0$. The correct map is found to be expressed by the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -f & 0 & 1 \end{pmatrix}$$

It may be seen that if $af << 1$ then the point $(a,b,1)^{\top}$ is mapped (almost) to itself by this transform.

5 Scene Reconstruction

We assume that the images have been transformed, point matches have been made and it is now required that the 3D scene be reconstructed point by point. Suppose a point $\mathbf{x} = (x,y,z,t)^{\top}$ is seen at locations $\mathbf{u} = (u,v,1)^{\top}$ in the first transformed image and at $\mathbf{u}' = (u+\delta,v,1)^{\top}$ in the second resampled image. This is the case, since disparities are parallel to the x-axis. It is desired to compute the coordinates of the point $\mathbf{x}$. As stated in Theorem 2.6 it is only possible to reconstruct the point up to a projective transformation of $\mathcal{P}^3$. As shown in [4], a first step in the reconstruction of 3D points up to perspective transformation of $\mathcal{P}^3$is to find a realization of the essential matrix Q. In the present case, this is particularly simple. According to Theorem 4.10, the essential matrix for the transformed images is or a particularly simple form, $Q = [(1,0,0)^{\top}]_{\times}$. A realization of Q is given by Theorem 3.7. The essential matrix is realized by the camera pair

$$P = (I \mid 0) \quad \text{and} \quad P' = (I \mid (1,0,0)^{\top}) \tag{9}$$

It is easily verified that the point

$$\mathbf{x} = (u, v, 1, \delta)^\top \tag{10}$$

is the 3D point mapping onto $\mathbf{u}$ and $\mathbf{u}'$. This makes the reconstruction of the scene almost trivial.

Looking closely at the form of the form of the reconstructed point (10) shows a curious effect of 3D projective transformation. If the disparity δ is zero, then the point $\mathbf{x}$ will be reconstructed as a point at infinity. Furthermore, choosing the M matrix as in section 3.2 will result in points with both positive and negative disparities. As δ changes from negative to positive, the reconstructed point will flip from near infinity in one direction to near infinity in the other direction. In other words, the reconstructed scene will straddle the plane at infinity, and if interpreted in a Euclidean sense by dehomogenization will contain points in diametrically opposite directions. This is all perfectly normal in the context of perspective geometry, but is a little disconcerting if one is accustomed to think in terms of Euclidean space. In the context of absolute stereo reconstruction, it causes no problem, since the process of tying down to an absolute Euclidean frame by the use of ground control points as described in [4] is done by selecting an appropriate normalizing perspective transformation.

We may observe here, however, that a different reconstruction of the 3D scene is possible which avoids points at infinity and diametrically splitting the scene. In particular, for any number α, it may be verified using Proposition 2.3 that the pair (P, P') where

$$P = \left(\begin{array}{ccc|c} 1 & 0 & \alpha & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right)$$

and $P' = (I \mid (1, 0, 0)^\top)$ is a realization of Q. In this case, the point

$$\mathbf{x} = (u - \alpha, v, 1, \delta + \alpha)^\top \tag{11}$$

is required reconstructed 3D point. By suitable choice of α it may be ensured that $\delta + \alpha > 0$ for all points.

6 Experimental results

The method was used to transform a pair of images of the Malibu area. Two images taken from widely different relatively oblique viewing angles are shown in Figures 1 and 2. A set of about 25 matched points were selected by hand and used to compute the essential matrix and epipolar transformation matrix.[1] The two 2D projective transformations necessary to transform them to matched epipolar projections were computed and applied to the images. Because of the great simplicity of the transformation, the resampling of the images may be done extremely quickly. With carefully programming, it is possible to resample the images in about 20 seconds each for 1024×1024 images on a Spark station 1A. The resulting resampled images are shown in Figures 3 and 4, which are placed side by side. As may be discerned, any disparities between the two images are parallel with the x-axis. By crossing the eyes it is possible to view the two images in stereo. The perceived scene looks a little strange, since it has undergone an apparent 3D transformation. However, the effect is not excessive.

[1] We also have the capability of selecting matched points automatically, despite the difference in viewpoint, but this capability was not used in this particular experiment

7 Conclusion

This paper gives a firm mathematical basis for the treatment of stereo images taken from widely different viewpoints. The method given avoids the necessity for camera calibration and provides significant gains in speed and ease of point matching. In addition, it makes the computational of the scene geometry extremely simple. The time taken to resample the image is negligeable compared with other processing time.

References

[1] Faugeras, O., "What can be seen in three dimensions with an uncalibrated stereo rig?", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 563 – 578.

[2] Faugeras, O. and Maybank, S., "Motion from Point Matches : Multiplicity of Solutions," International Journal of Computer Vision, **4**, (1990), 225–246.

[3] H.C. Longuet-Higgins, *"A computer algorithm for reconstructing a scene from two projections,"* Nature, Vol. 293, 10 Sept. 1981.

[4] R. Hartley, R. Gupta and T. Chang, "Stereo from Uncalibrated Cameras", Proceedings Computer Vision and Pattern Recognition conference (CVPR-92), 1992.

[5] R. Hartley, "Estimation of Relative Camera Positions for Uncalibrated Cameras,", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 579 – 587.

[6] I.E. Sutherland, *"Three dimensional data input by tablet,"* Proceedings of IEEE, Vol. 62, No. 4, pp. 453–461, April 1974.

Appendix : Analysis of constraints

The matrix of coefficients in the set of equations (5) may written in the form

$$\begin{pmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_3 \\ Q_2 & -Q_1 & 0 \\ -Q_3 & 0 & Q_1 \\ 0 & Q_3 & -Q_2 \end{pmatrix} \tag{12}$$

where Q_i represents the three entries $q_{1i}q_{2i}q_{3i}$ making up the i-th column of the matrix Q. First, we consider the case where $Q_1 = 0$. Since Q has rank 2, it can not be the case that Q_2 is a multiple of Q_3. In this case, the first row of the matrix is zero, but the remaining five rows are linearly independent. The same argument holds if $Q_2 = 0$ or $Q_3 = 0$.

Next, consider the case where $Q_2 = \alpha Q_1$. Since Q has rank 2, it can not be that $Q_3 = \beta Q_1$ or $Q_3 = \beta Q_2$. In this case, the fourth row of matrix (12) is dependent on the first two rows, but the other five rows are linearly independent.

Finally, consider the case where no column of Q is a simple multiple of another row, but $\alpha Q_1 + \beta Q_2 + \gamma Q_3 = 0$. Then, it may be verified that multiplying the rows of the matrix by factors $\alpha, -\beta^2/\alpha, -\gamma^2/\alpha, \beta, -\gamma, -\beta\gamma/\alpha$ and adding results in 0. In other words, the rows are linearly dependent. On the other hand, it can be shown by a straight-forward argument that any five of the rows are linearly independent.

Thus, in general, we have a set of 6 restraints on the entries of M, only 5 of which are linearly independent.

Figure 1. First view of Malibu

Figure 2. Second view of Malibu

Figure 3. First Resampled view of Malibu

Figure 4. Second resampled view of Malibu