Theory and Practice of Projective Rectification ⋆

Richard I. Hartley

G.E. CRD, 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 fundamental matrix of Longuet-Higgins which describes the epipolar geometry of the image pair. The approach taken is consistent with that advocated by Faugeras ([4]) of avoiding camera calibration. The paper uses methods of projective geometry to determine a pair of 2D projective transformations 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

An approach to stereo reconstruction that avoids the necessity for camera calibration was described in [7, 4] In those papers it was shown that the the 3-dimensional configuration of a set of points is determined up to a projectivity of the 3-dimensional projective space $\mathcal{P}^3$by their configuration in two independent views from uncalibrated cameras. 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 [7] it is shown that the *fundamental matrix*, F, ([10]) 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 previous papers start 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 in matching images taken from very different locations, perspective distortion and different viewpoint make corresponding regions look very different. The image rectification method described here overcomes this problem by

⋆The research described in this paper has been supported by DARPA Contract #MDA972-91-C-0053

transforming both images to a common reference frame. It may be used as a preliminary step to comprehensive image matching, greatly simplifying the image matching problem. The approach taken is consistent with the projective-geometrical methods used in [4] and [7].

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 rig. For an arbitrary placement of cameras, however, the epipolar geometry will be more complex. In effect, transforming the two images by the appropriate projective transformations reduces the problem to that of a rectilinear stereo rig. Many stereo algorithms described in previous literature have assumed a rectilinear or near-rectilinear stereo rig.

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 transformations which subject the images to minimal distortion.

The advantages of reducing to the case of a rectilinear stereo rig 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 makes use of the fundamental matrix F. The scene may be reconstructed up to a 3D projectivity from the resampled images. Because we are effectively dealing with a rectilinear stereo rig, the mathematics of this reconstruction is extremely simple. In fact, once the two images have been transformed, the original images may be thrown away and the transformations forgotten, since unless parametrized camera models are to be computed (which we wish to avoid), the resampled images are as good as the original ones. If ground control points, or other constraints on the scene are known, it is possible to compute the absolute (Euclidean) configuration of the scene from the projective reconstruction ([7]).

A lengthy discussion of other methods of rectification is given in the Manual of Photogrammetry ([11]), including a description of graphical, optical and software techniques. Optical techniques have been widely used in the past, but are being replaced by software methods that model the geometry of optical projection. Notable among these latter is the algorithm of Ayache et. al ([1, 2]) which uses knowledge of the camera matrices to compute a pair of rectifying transformations. They also give a manner of rectifying a triple of images using both horizontal and vertical epipolar lines on one of the images. In contrast with their algorithm, the present method does not need the camera matrices, but relies on point correspondences alone. An additional feature of the algorithm described in this paper is that it minimizes the horizontal disparity of points along the epipolar lines so as to minimize the range of search for further matched points.

Other notable recent papers include [15] which deals with rectification of sequences used for rover navigation, and [13] which, however, considers only the special case of partially aligned cameras. In addition, [19, 3, 17, 16] use rectification for various special purpose imaging situations.

1.1 Preliminaries

Column vectors will be denoted by bold lower-case letters, such as $\mathbf{x}$. Row vectors are transposed column vectors, such as $\mathbf{x}^\mathrm{T}$. Thus, the inner product of two vectors is represented by $\mathbf{a}^\mathrm{T}\mathbf{b}$. On the other hand, $\mathbf{a}\mathbf{b}^\mathrm{T}$ is a matrix of rank 1. Matrices will be denoted by upper case letters. The notation $\approx$ is used to indicate equality of vectors or matrices up to multiplication by a non-zero 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 $A^* \approx (A^\mathrm{T})^{-1}$.

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

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

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 $\mathbf{t}^\mathrm{T}[\mathbf{t}]_\times = [\mathbf{t}]_\times\mathbf{t} = 0$ 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 $\mathbf{s}^\mathrm{T}[\mathbf{t}]_\times = \mathbf{s} \times \mathbf{t}$ 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* $\mathbf{t}$

$$M^*[\mathbf{t}]_\times = [M\mathbf{t}]_\times M \tag{2}$$

Projective Geometry : Real projective n-space consists of the set of equivalence classes of non-zero real $(n+1)$-vectors, where two vectors are considered equivalent if they differ by a constant factor. A vector representing a point in $\mathcal{P}^n$ in this way is known as a *homogeneous coordinate* representation of the point.

Real projective n-space contains Euclidean n-space as the set of all homogeneous vectors with final coordinate not equal to zero. For example, a point in $\mathcal{P}^2$ is represented by a vector $\mathbf{u} = (u, v, w)^\mathrm{T}$. If $w \neq 0$, then this represents the point in R^2 expressed in Euclidean coordinates as $(u/w, v/w)^\mathrm{T}$.

Lines in $\mathcal{P}^2$ are also represented in homogeneous coordinates. In particular, the line $\boldsymbol{\lambda}$ with coordinates $(\lambda, \mu, \nu)^\mathrm{T}$ is the line consisting of points satisfying equation $\lambda u + \mu v + \nu w = 0$. In other words, a point $\mathbf{u}$ lies on a line $\boldsymbol{\lambda}$ if and only if $\boldsymbol{\lambda}^\mathrm{T}\mathbf{u} = 0$. The line joining two points $\mathbf{u}_1$ and $\mathbf{u}_2$ is given by the cross product $\mathbf{u}_1 \times \mathbf{u}_2$.

A projective mapping (or transformation) from $\mathcal{P}^n$ to $\mathcal{P}^m$ is a map represented by a linear transformation of homogeneous coordinates. Projective mappings may be represented by matrices of dimension $(m+1)\times(n+1)$. The word projectivity will also be used to denote an invertible projective mapping.

If A is a 3×3 non-singular matrix representing a projective transformation of $\mathcal{P}^2$, then A^* is the corresponding line map. In other words, if $\mathbf{u}_1$ and $\mathbf{u}_2$ line on a line $\boldsymbol{\lambda}$, then $A\mathbf{u}_1$ and $A\mathbf{u}_2$ line on the line $A^*\boldsymbol{\lambda}$: in symbols

$$A^*(\mathbf{u}_1\times\mathbf{u}_2) = (A\mathbf{u}_1)\times(A\mathbf{u}_2) \quad .$$

This formula is easily derived from Proposition 1.1.

Most of the vectors and matrices used in this paper will be defined only up to multiplication by a nonzero factor. Usually, we will ignore multiplicative factors and use the equality sign (=) to denote equality up to a constant factor. Exceptions to this rule will be noted specifically.

Camera Model. The camera model considered in this paper is that of central projection, otherwise known as the pinhole or perspective camera model. Such a camera maps a region of R^3 lying in front of the camera into a region of the image plane R^2. For mathematical convenience we extend this mapping to a mapping between projective spaces $\mathcal{P}^3$ (the object space) and $\mathcal{P}^2$ (image space). The map is defined everywhere in $\mathcal{P}^3$ except at the centre of projection of the camera (or *camera centre*).

Points in object space will therefore be denoted by homogeneous 4-vectors $\mathbf{x} = (x, y, z, t)^{\mathrm{T}}$, or more usually as $(x, y, z, 1)^{\mathrm{T}}$. Image space points will be represented by $\mathbf{u} = (u, v, w)^{\mathrm{T}}$.

The projection from object to image space is a projective mapping represented by a 3×4 matrix P of rank 3, known as the *camera matrix*. The camera matrix transforms points in 3-dimensional object space to points in 2-dimensional image 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.

Suppose the camera centre is not at infinity, and let its Euclidean coordinates be $\mathbf{t} = (t_x, t_y, t_z)^{\mathrm{T}}$. The camera mapping is undefined at $\mathbf{t}$ in that $P(t_x, t_y, t_z, 1)^{\mathrm{T}} = 0$. If P is written in block form as $P = (M \mid \mathbf{v})$, then it follows that $M\mathbf{t}+\mathbf{v} = 0$, and so $\mathbf{v} = -M\mathbf{t}$. Thus, the camera matrix may be written in the form

$$P = (M \mid -M\mathbf{t})$$

where $\mathbf{t}$ is the camera centre. Since P has rank 3, it follows that M is non-singular.

2 Epipolar Geometry

Suppose that we have two images of a common scene and let $\mathbf{u}$ be a point in the first image. The locus of all points in $\mathcal{P}^3$ that map to $\mathbf{u}$ consists of a straight line through the centre of the first camera. As seen from the second camera this straight line maps to a straight line in the image known as a *epipolar line*. Any point $\mathbf{u}'$ in the second image matching point $\mathbf{u}$ must lie on this epipolar line. The epipolar lines in the second image corresponding to 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.

Thus, there exists a mapping from points in the first image to epipolar lines in the second image. It is a basic fact that this mapping is a projective mapping. In particular, there exists ([10, 5, 8]) a 3×3 matrix F called the *fundamental matrix* which maps points in the first image to the corresponding epipolar line in the second image according to the mapping $\mathbf{u} \mapsto F\mathbf{u}$. 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 $F\mathbf{u}_i$ means that

$$\mathbf{u}'^{\mathrm{T}}_i F \mathbf{u}_i = 0 \ . \tag{3}$$

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

The following theorem gives some basic well known properties of the fundamental matrix.

Proposition 2.2. *Suppose that F is the fundamental matrix corresponding to an ordered pair of images (J, J') and $\mathbf{p}$ and $\mathbf{p}'$ are the epipoles.*

1. *Matrix F^{T} is the fundamental matrix corresponding to the ordered pair of images (J', J).*
2. *F factors as a product $F = [\mathbf{p}']_\times M = M^*[\mathbf{p}]_\times$ for some non-singular matrix M.*
3. *The epipole $\mathbf{p}$ is the unique point such that $F\mathbf{p} = 0$. Similarly, $\mathbf{p}'$ is the unique point such that $\mathbf{p}'^{\mathrm{T}} F = 0$.*

According to Proposition 2.2, the matrix F determines the epipoles in both images. Furthermore, F provides the map between points in one image and epipolar lines in the other image. Thus, the complete geometry and correspondence of epipolar lines is encapsulated in the fundamental matrix.

The fact that F factorizes into a product of non-singular and skew-symmetric matrices is a basic property of the fundamental matrix. The factorization is not unique, however, as is shown by the following proposition.

Proposition 2.3. *Let the 3×3 matrix F factor in two different ways as $F = S_1 M_1 = S_2 M_2$ where each S_i is a non-zero skew-symmetric matrix and each M_i is non-singular. Then $S_2 = S_1$. Furthermore, if $S_i = [\mathbf{p}']_\times$ then $M_2 = (I + \mathbf{p}'\mathbf{a}^{\mathrm{T}})M_1$ for some vector $\mathbf{a}$.*

Conversely, if $M_2 = (I + \mathbf{p}'\mathbf{a}^{\mathrm{T}})M_1$, then $[\mathbf{p}']_\times M_1 = [\mathbf{p}']_\times M_2$.

Proof. If $\mathbf{p}'^{\mathrm{T}} F = 0$, then it follows that $S_1 = S_2 = [\mathbf{p}']_\times$. This proves the first claim. Now suppose that $[\mathbf{p}']_\times M_1 = [\mathbf{p}']_\times M_2$. Then $[\mathbf{p}']_\times = [\mathbf{p}']_\times M_2 M_1^{-1}$ and so $[\mathbf{p}']_\times (M_2 M_1^{-1} - I) = 0$. It follows that each column of $M_2 M_1^{-1} - I$ is a scalar multiple of $\mathbf{p}'$. Therefore $M_2 M_1^{-1} - I = \mathbf{p}'\mathbf{a}^{\mathrm{T}}$ for some vector $\mathbf{a}$. Hence $M_2 = (I + \mathbf{p}'\mathbf{a}^{\mathrm{T}})M_1$ as desired.

The converse may be verified very easily, using the fact that $[\mathbf{p}']_\times \mathbf{p}' = 0$. □

As seen, the fundamental matrix provides a mapping between points in one image and the corresponding epipolar lines in the other image. We now ask which (point-to-point) projective transformations from image J to image J' take epipolar lines to corresponding epipolar lines. Such a transformation will be said to "preserve epipolar lines." The question is completely answered by the following result which characterizes such mappings.

Proposition 2.4. *Let F be a fundamental matrix and $\mathbf{p}$ and $\mathbf{p}'$ the two epipoles. If F factors as a product $F = [\mathbf{p}']_\times M$ then*

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

2. *If $\mathbf{u}$ is a point in the first image, then $M\mathbf{u}$ lies on the corresponding epipolar line $F\mathbf{u}$ in the second image.*

3. *If $\boldsymbol{\lambda}$ is a line in the first image, passing through the epipole $\mathbf{p}$ (that is, an epipolar line), then $M^*\boldsymbol{\lambda}$ is the corresponding epipolar line in the other image.*

Conversely, if M is any matrix satisfying condition 2, or 3, then F factors as a product $F = [\mathbf{p}']_\times M$.

Proof.
1. $F = [\mathbf{p}']_\times M$, so $0 = F\mathbf{p} = [\mathbf{p}']_\times M\mathbf{p} = \mathbf{p}' \times (M\mathbf{p})$. It follows that $M\mathbf{p} = \mathbf{p}'$.
2. The condition that $M\mathbf{u}$ lies on $F\mathbf{u}$ for all $\mathbf{u}$ is the same as saying that the epipolar line $\mathbf{p}' \times M\mathbf{u}$ equals $F\mathbf{u}$ for all $\mathbf{u}$. This is equivalent to the condition $[\mathbf{p}']_\times M = F$, as required.
3. Let $\boldsymbol{\lambda}$ be an epipolar line in the first image and let $\mathbf{u}$ be a point on it. Then

$$\boldsymbol{\lambda}' = F\mathbf{u} = M^*[\mathbf{p}]_\times\mathbf{u} = M^*(\mathbf{p} \times \mathbf{u}) = M^*\boldsymbol{\lambda}$$

as required. □

Condition 2 in the above theorem simply states that M preserves epipolar lines. Thus, the projective point maps from J to J' that preserve epipolar lines are precisely those represented by matrices M appearing in a factorization of F. Since the factorization of F is not unique, however (see Proposition 2.3), there exists a 3-parameter family of such transformations.

3 Mapping the Epipole to Infinity

In this section we will discuss the question of finding a projective transformation H of an image mapping an epipole to a point at infinity. In fact, if epipolar lines are to be transformed to lines parallel with the x axis, then the epipole should be mapped to the infinite point $(1, 0, 0)^{\mathrm{T}}$. This leaves many degrees of freedom (in fact four) open for H, and if an inappropriate H is chosen, severe projective distortion of the image can take place. In order that the resampled image should look somewhat like the original image, we may put closer restrictions on the choice of H.

One condition that leads to good results is to insist that the transformation H should act as far as possible as a rigid transformation in the neighbourhood of a given selected point $\mathbf{u}_0$ of the image. By this is meant that to first order 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 the context of aerial photography if the view is known not to be excessively oblique.

For the present, suppose $\mathbf{u}_0$ is the origin and the epipole $\mathbf{p} = (f, 0, 1)$ lies on the x axis. Now consider the following transformation.

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{pmatrix} \tag{4}$$

This transformation takes the epipole $(f, 0, 1)^{\mathrm{T}}$ to the point at infinity $(f, 0, 0)^{\mathrm{T}}$ as required. A point $(u, v, 1)^{\mathrm{T}}$ is mapped by G to the point $(\hat{u}, \hat{v}, 1)^{\mathrm{T}} = (u, v, 1 - u/f)^{\mathrm{T}}$. If $|u/f| < 1$ then we may write

$$(\hat{u}, \hat{v}, 1)^{\mathrm{T}} = (u, v, 1 - u/f) = (u(1 + u/f + \ldots), v(1 + u/f + \ldots), 1)^{\mathrm{T}} \ .$$

The Jacobian is

$$\frac{\partial(\hat{u}, \hat{v})}{\partial(u, v)} = \begin{pmatrix} 1 + 2u/f & 0 \\ v/f & 1 + u/f \end{pmatrix}$$

plus higher order terms in u and v. Now if $u = v = 0$ then this is the identity map. In other words, G is approximated (to first order) at the origin by the identity mapping.

For an arbitrarily placed point of interest $\mathbf{u}_0$ and epipole $\mathbf{p}$, the required mapping H is a product $H = GRT$ where T is a translation taking the point $\mathbf{u}_0$ to the origin, R is a rotation about the origin taking the epipole $\mathbf{p}'$ to a point $(f, 0, 1)^{\mathrm{T}}$ on the x axis, and G is the mapping just considered taking $(f, 0, 1)^{\mathrm{T}}$ to infinity. The composite mapping is to first order a rigid transformation in the neighbourhood of $\mathbf{u}_0$.

4 Matching Transformations

We consider two images J and J'. The intention is to resample these two images according to transformations H to be applied to J and H' to be applied to J'. The resampling is to be done in such a way that an epipolar line in J is matched with its corresponding epipolar line in J'. More specifically, if $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are any pair of corresponding epipolar lines in the two images, then $H^*\boldsymbol{\lambda} = H'^*\boldsymbol{\lambda}'$. (Recall that H^* is the line map corresponding to the point map H.) Any pair of transformations satisfying this condition will be called a *matched pair* of transformations.

Our strategy in choosing a matched pair of transformations is to choose H' first to be some transformation that sends the epipole $\mathbf{p}'$ to infinity as described in the previous section. We then seek a matching transformation H chosen so as to minimize the sum-of-squares distance

$$\sum_i d(H\mathbf{u}_i, H'\mathbf{u}'_i)^2 \ . \tag{5}$$

The first question to be determined is how to find a transformation matching H'. That question is answered in the following theorem.

Theorem 4.5. *Let J and J' be images with fundamental matrix $F = [\mathbf{p}']_\times M$, and let H' be a projective transformation of J'. A projective transformation H of J matches H'*

if and only if H is of the form

$$H = (I + H'\mathbf{p}'\mathbf{a}^{\mathrm{T}})H'M \tag{6}$$

for some vector $\mathbf{a}$.

Proof. If $\mathbf{u}$ is a point in J, then $\mathbf{p} \times \mathbf{u}$ is the epipolar line in the first image, and $F\mathbf{u}$ is the epipolar line in the second image. Transformations H and H' are a matching pair if and only if $H^*(\mathbf{p} \times \mathbf{u}) = H'^*F\mathbf{u}$. Since this must hold for all $\mathbf{u}$ we may write equivalently $H^*[\mathbf{p}]_\times = H'^*F = H'^*[\mathbf{p}']_\times M$ or, applying Proposition 1.1,

$$[H\mathbf{p}]_\times H = [H'\mathbf{p}']_\times H'M \tag{7}$$

which is a necessary and sufficient condition for H and H' to match. In view of Proposition 2.3, this implies (6) as required.

To prove the converse, if (6) holds, then

$$\begin{aligned} H\mathbf{p} &= (I + H'\mathbf{p}'\mathbf{a}^{\mathrm{T}})H'M\mathbf{p} \\ &= (I + H'\mathbf{p}'\mathbf{a}^{\mathrm{T}})H'\mathbf{p}' \\ &= (I + \mathbf{a}^{\mathrm{T}}H'\mathbf{p}')H'\mathbf{p}' \\ &\approx H'\mathbf{p}' \ . \end{aligned}$$

According to Proposition 2.3, this, along with (6) are sufficient for (7) to hold, and so H and H' are matching transforms. □

We are particularly interested in the case when H' is a transformation taking the epipole $\mathbf{p}'$ to a point at infinity $(1, 0, 0)^{\mathrm{T}}$. In this case, $I + H'\mathbf{p}'\mathbf{a}^{\mathrm{T}} = I + (1, 0, 0)^{\mathrm{T}}\mathbf{a}^{\mathrm{T}}$ is of the form

$$A = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

which represents an affine transformation. Thus, a special case of Theorem 4.5 is

Corollary 4.6. *Let J and J' be images with fundamental matrix $F = [\mathbf{p}']_\times M$, and let H' be a projective transformation of J' mapping the epipole $\mathbf{p}'$ to the infinite point $(1, 0, 0)^{\mathrm{T}}$. A transform H of J matches H' if and only if H is of the form $H = AH_0$, where $H_0 = H'M$ and A is an affine transformation of the form (8).*

Given H' mapping the epipole to infinity, we may use this corollary to make the choice of a matching transformation H to minimize the disparity. Writing $\hat{\mathbf{u}}'_i = H'\mathbf{u}'_i$ and $\hat{\mathbf{u}}_i = H_0\mathbf{u}_i$, the minimization problem (5) is to find A of the form (8) such that

$$\sum_i d(A\hat{\mathbf{u}}_i, \hat{\mathbf{u}}'_i)^2 \tag{9}$$

is minimized.

In particular, let $\hat{\mathbf{u}}_i = (\hat{u}_i, \hat{v}_i, 1)$, and let $\hat{\mathbf{u}}'_i = (\hat{u}'_i, \hat{v}_i, 1)$. Since H' and M are known, these vectors may be computed from the matched points $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$. Then the quantity to be minimized (9) may be written as

$$\sum_i (a\hat{u}_i + b\hat{v}_i + c - \hat{u}'_i)^2 + (\hat{v}_i - \hat{v}'_i)^2 \ .$$

Since $(\hat{v}_i - \hat{v}'_i)^2$ is a constant, this is equivalent to minimizing

$$\sum_i (a\hat{u}_i + b\hat{v}_i + c - \hat{u}'_i)^2 \ .$$

This is a simple linear least-squares parameter minimization problem, and is easily solved using known linear techniques ([14]) to find a, b and c. Then A is computed from (8) and H from (6). Note that a linear solution is possible because A is an affine transformation. If it were simply a projective transformation, this would not be a linear problem.

5 Quasi-affine Transformations

In resampling an image via a 2D projective transformation, it is possible to split the image so that connected regions are no longer connected. Such a projective transformation is called a non-quasi-affine projectivity. (A more precise definition is given below.) An example is given in Figure 1 which show an image of a comb and the image resampled according to a non-quasi-affine projectivity. Resampling an image via such a projectivity is obviously undesirable. Therefore we will consider methods of avoiding such cases.

In real images, the whole of the image plane is not visible, but usually only some rectangular region of the image plane. Accordingly, we introduce the concept of a *view window*. The view window is the part of the image plane that contain all available imagery, including matched points (but not necessarily the epipole). In resampling an image, only points in the view window will be resampled. The view window is assumed to be a convex subset of the image plane.

The line at infinity L_∞ in the projective plane $\mathcal{P}^2$ consists of all points with final coordinate equal to 0. Let W be a convex region of the plane. A projective transformation H is said to be *quasi-affine* with respect to W if $H(W) \cap L_\infty = \emptyset$. It is clear that if the epipole $\mathbf{p}$ lies outside of the convex view-window W, then there exists a projectivity, quasi-affine with respect to W, taking $\mathbf{p}$ to $(0, 0, 1)^{\mathrm{T}}$. In fact, any line through $\mathbf{p}$ not meeting W may be chosen as the line $H^{-1}(L_\infty)$. The perspectivity (4) maps the line $u = f$ to infinity – that is, the line through the epipole parallel to the vertical image axis. If the epipole lies sufficiently far away from the view window, then this mapping will be quasi-affine.

If the epipole lies inside the view window of an image, the techniques of this paper may still be applied by considering a smaller view window. It is possible that the projectivity H' constructed in Section 6 is not quasi-affine, in which case the view window should be shrunk, or some other projectivity chosen.

We now turn to the question of when it is possible to find a pair of matched projectivities H and H' each one quasi-affine with respect to the view window of the respective image. It is not to be expected that this will always be possible even when the epipoles lie outside of both view windows. However, one can do almost as well, as the following theorem shows.

Theorem 5.7. *Consider images J and J' with view windows W and W'. Suppose the epipole $\mathbf{p}'$ in image J' does not lie in W'. Let H' be a projectivity of J', quasi-affine with respect to W', and mapping $\mathbf{p}'$ to infinity, and let H be any matching projectivity. Then there exists a convex subwindow $W_+ \subseteq W$ such that H is quasi-affine with respect to W_+ and such that W_+ contains all points in W that match a point in W'.*

If our purpose in resampling is to facilitate point matching, then W_+ contains all the interesting part of the image J. The theorem asserts that H is quasi-affine with respect to W_+, and so W_+ may be resampled according to H with satisfactory results. Before providing the proof of Theorem 5.7 some preliminary material is necessary.

A set of image correspondences is called a *defining set* of image correspondences if equations (3) have a unique solution. Thus, a defining set of correspondences is one that contains sufficiently many matches to determine the epipolar structure of the image pair.

Given a set of image correspondences $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$. Let P and P' be camera matrices and $\{\mathbf{x}_i\}$ be a set of points in $\mathcal{P}^3$. The triple $(P, P', \{\mathbf{x}_i\})$ is called a realization of the set of correspondences $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$ if $\mathbf{u}_i = P\mathbf{x}_i$ and $\mathbf{u}'_i = P'\mathbf{x}_i$. Note that it does not follow that P and P' are the actual camera matrices, or that the points $\mathbf{x}_i$ represent the actual point locations in space. Indeed, without camera calibration, the actual structure of the point set $\{\mathbf{x}_i\}$ may be determined from image matches **only** up to a 3D projective transformation ([7, 4]).

The following theorems are based on an analysis of *cheirality*, that is, determination of which points are behind and which points are in front of a camera. Proofs are given in [9], which contains a thorough investigation of cheirality.

Note that when equality of vectors (=) is considered in the following theorems, we mean exact equality, and not equality up to a factor.

Theorem 5.8. *Let H be a 2D projectivity and $H(u_i, v_i, 1)^{\mathrm{T}} = \alpha_i(\hat{u}_i, \hat{v}_i, 1)^{\mathrm{T}}$ for some set of points $\mathbf{u}_i = (u_i, v_i, 1)^{\mathrm{T}} \in R^2$. Then H is quasi-affine with respect to $\{\mathbf{u}_i\}$ if and only if all α_i have the same sign.*

Theorem 5.9. *Let $(P, P', \{\mathbf{x}_i\})$ be a realization of a defining set of image correspondences $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$ derived from a physically realizable 3D point set. Let $\mathbf{x}_i \approx (x_i, y_i, z_i, t_i)^{\mathrm{T}}$, $\mathbf{u}_i \approx (u_i, v_i, w_i)^{\mathrm{T}}$ and $(u_i, v_i, w_i)^{\mathrm{T}} = P(x_i, y_i, z_i, t_i)^{\mathrm{T}}$. Let primed quantities be similarly defined. Then the sign of $w_i w'_i$ is constant for all i.*

Now, we can prove Theorem 5.7.

Proof. Consider matching projectivities H and H' for a pair of images J and J', and suppose H' is quasi-affine with respect to W'. Let $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$ be a defining set of correspondences with $\mathbf{u}_i = (u_i, v_i, 1)^{\mathrm{T}} \in W$ and $\mathbf{u}'_i = (u'_i, v'_i, 1)^{\mathrm{T}} \in W'$ for all i. Let $H(u_i, v_i, 1)^{\mathrm{T}} = \alpha_i(\hat{u}_i, \hat{v}_i, 1)^{\mathrm{T}}$ and $H'(u'_i, v'_i, 1)^{\mathrm{T}} = \alpha'_i(\hat{u}_i + \delta_i, \hat{v}_i, 1)^{\mathrm{T}}$. Since H' is quasi-affine, all α'_i have the same sign. We wish to prove that all α_i have the same sign.

A realization for $\mathbf{u}'_i \leftrightarrow \mathbf{u}_i$ is given by setting $P = H^{-1}(I \mid 0)$, $P' = H'^{-1}(I \mid (1, 0, 0)^{\mathrm{T}})$ and $\mathbf{x}_i = (\hat{u}_i, \hat{v}_i, 1, \delta_i)^{\mathrm{T}}$. Indeed we verify that

$$P\mathbf{x}_i = H^{-1}(I \mid 0)(\hat{u}_i, \hat{v}_i, 1, \delta_i)^{\mathrm{T}} = H^{-1}(\hat{u}_i, \hat{v}_i, 1)^{\mathrm{T}} = \alpha_i^{-1}(u_i, v_i, 1)^{\mathrm{T}}$$

and

$$P'\mathbf{x}_i = H'^{-1}(I \mid (1, 0, 0)^{\mathrm{T}})(\hat{u}_i, \hat{v}_i, 1, \delta_i)^{\mathrm{T}} = H'^{-1}(\hat{u}_i + \delta_i, \hat{v}_i, 1)^{\mathrm{T}} = \alpha_i'^{-1}(u'_i, v'_i, 1)^{\mathrm{T}} \ .$$

It follows from Theorem 5.9 that $(\alpha_i \alpha'_i)^{-1}$ and hence $\alpha_i \alpha'_i$ has the same sign for all i. However, by hypothesis, H' is quasi-affine with respect to all $\mathbf{u}'_i$, and so all α'_i have the same sign, and therefore so do all α_i. This means that all $\mathbf{u}_i$ lie to one side of $H^{-1}(L_\infty)$. We define W_+ to be the part of W lying to the same side of $H^{-1}(L_\infty)$ as all $\mathbf{u}_i$, and the proof is complete. □

6 Resampling

Once the two resampling transformations H and H' have been determined, the pair of images may be resampled. There are two steps, first determine the extent of the resampled images, and second, carry out the resampling.

6.1 Determining the Dimensions of the Output Image

Assume that the range of the resampling projectivities H and H' are the same plane $\hat{J}$. Suppose that the projectivity H' is quasi-affine with respect to the window, W'. Then $H'(W)$ will be a convex region in $\hat{J}$. In fact, if W is a polygonal region, then so is $H'(W')$, and it may be easily computed. As for $H(W)$, if H is quasi-affine with respect to W, then $H(W)$ will also be a bounded convex set. On the other hand, if H is not quasi-affine with respect to the whole window W, then $H(W)$ will split into two parts, $H(W_+)$ and $H(W_-)$ stretching to infinity, along with points at infinity in $\hat{J}$. According to Theorem 5.7, only one of the two regions W_- and W_+ (let us suppose W_+) contains points that match points in W'. It is a matter of straight-forward geometrical programming to determine the intersection $H'(W') \cap H(W_+)$. The resampling window $\hat{W}$ in $\hat{J}$ may then be chosen. Normally, $\hat{W}$ should be a rectangular region aligned with the coordinate axes in $\hat{J}$. It is a matter of choice whether $\hat{W}$ is chosen as the smallest rectangle containing $H'(W') \cap H(W_+)$ or whether $\hat{W}$ should be a rectangle contained in $H'(W') \cap H(W_+)$.

6.2 Resampling

Once the window $\hat{W}$ is chosen, it is a simple matter to resample each of the images. Consider resampling the first image. For each pixel location (that is, point with integer coordinates) $\hat{\mathbf{u}}$ in $\hat{W}$, the corresponding location $\mathbf{u} = H^{-1}\hat{\mathbf{u}}$ in J is computed, and the "colour" or image intensity at $\mathbf{u}$ is determined. Pixel $\hat{\mathbf{u}}$ is then coloured with this colour. If $\mathbf{u}$ lies outside the view window W, or W_+, then pixel $\hat{\mathbf{u}}$ is coloured some chosen background color. Since point $\mathbf{u}$ will not have integer coordinates, it is necessary in determining the colour of $\mathbf{u}$ to interpolate. In the images shown in this paper, linear interpolation was used, and gave adequate results. In other cases, such as if aliasing becomes an important issue, some more sophisticated method of interpolation should be used ([20]).

7 Scene Reconstruction

We assume that the images have been resampled, point matches have been made and it is required that the 3D scene be reconstructed point by point. Without knowledge of camera parameters, the scene can be reconstructed up to a 3D projectivity. Suppose a point $\mathbf{x}_i$ is seen at locations $\mathbf{u}_i = (u_i, v_i, 1)^{\mathrm{T}}$ in the first resampled image and at $\mathbf{u}'_i = (u_i + \delta_i, v_i, 1)^{\mathrm{T}}$ in the second resampled image. Note that disparities are parallel to the x-axis. It was seen in the proof of Theorem 5.7 that a possible reconstruction is with $P = (I|0)$, $P' = (I|(1,0,0)^{\mathrm{T}})$ and $\mathbf{x}_i = (u_i, v_i, 1, \delta_i)^{\mathrm{T}}$.

Looking closely at the form of the reconstructed point shows a curious effect of 3D projective transformation. If the disparity δ_i is zero, then the point $\mathbf{x}_i$ will be reconstructed as a point at infinity. As δ_i 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 will contain points in diametrically opposite directions.

A different reconstruction of the 3D scene is possible which avoids points at infinity and diametrically splitting the scene. In particular, if one of the images is translated by a distance α in the x direction, then a disparity of δ_i becomes a disparity of $\delta_i + \alpha$. It follows that the scene may be reconstructed with $\mathbf{x}_i = (u_i, v_i, 1, \delta_i + \alpha)^{\mathrm{T}}$ are positive, that is $\delta_i + \alpha > 0$ for all i. Note that the eye makes this adjustment automatically when viewing a stereo pair of images such as Figure 3, resulting in a sensible perception of the scene.

8 Algorithm Outline

The resampling algorithm will now be summarized. The input is a pair of images containing a common overlap region. The output is a pair of images resampled so that the epipolar lines in the two images are horizontal (parallel with the $\mathbf{u}$ axis), and such that corresponding points in the two images are as close to each other as possible. Any remaining disparity between matching points will be along the the horizontal epipolar lines. A top-level outline of the image is as follows.

1. Identify a seed set of image-to-image matches $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ between the two images. Seven points at least are needed, though more are preferable. It is possible to find such matches by automatic means.

2. Compute the fundamental matrix F and find the epipoles $\mathbf{p}$ and $\mathbf{p}'$ in the two images. The linear method of computation of F as the least-squares solution to equations (3) can be used, requiring eight point matches or more. For best results, the linear solution can be used as a basis for iteration to a least-squares solution.

3. Select a projective transformation H' that maps the epipole $\mathbf{p}'$ to the point at infinity, $(1, 0, 0)^{\mathrm{T}}$. The method of Section 3 gives good results.

4. Find the matching projective transformation H that minimizes the least-squares distance
$$\sum_i d(H\mathbf{u}_i, H'\mathbf{u}'_i) \ . \tag{10}$$
The method used is a linear method described in Section 4.

5. Resample the first image according to the projective transformation H and the second image according to the projective transformation H'. A simple method is given in Section 6.

9 Examples

9.1 A pair of aerial images

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 Figure 2. A set of about 25 matched points were selected by hand and used to compute the fundamental matrix. The two 2D projective transformations necessary to transform them to matched epipolar projections were computed and applied to the images. The resulting resampled images are shown side-by-side in Figure 3 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 projective transformation. However, the effect is not excessive.

9.2 A pair of images of an airfield

In the next example (see Figure 4), the the pair of images taken on different days are rectified to create a stereo pair. These two images may be merged to create a 3D impression. This example suggests that rectification may be used as an aid in detecting changes in the two images, which become readily apparent when the images are viewed stereoscopically.

9.3 Satellite images

Although satellite images are not strictly speaking perspective images, they are sufficiently close to being perspective projections that the techniques described in this paper may be applied. To demonstrate this, the algorithm was applied to two satellite images, one Landsat image and one SPOT image. These two images have different resolutions and different imaging geometries. In this case, a set of seed matches were found, this time automatically, and the two images were resampled. (The method of finding seed matches for differently oriented images is based on local resampling.) The results are shown in Figure 5 which shows the images after resampling.

9.4 Non-aerial images

The method may also be applied to images other than aerial images. Figure 6 shows a pair of images of some wooden block houses. Edges and vertices in these two images were extracted automatically and a small number of common vertices were matched by hand. The two images were then resampled according to the methods desribed here. The results are shown in Figure 7. In this case, because of the wide difference in viewpoint, and the three-dimensional shape of the objects, the two images even after resampling look quite different. However, it is the case that any point in the first image will now match a point in the second image with the same y coordinate. Therefore, in order to find further point matches between the images only a one-dimensional search is required.

10 Conclusion

This paper gives a firm mathematical basis as well as a rapid practical algorithm for the rectification 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. Because of the great simplicity of the projective 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 Sparc station 1A.

References

[1] Nicholas Ayache and Charles Hansen, "Rectification of Images for Binocular and Trinocular Stereovision", Proc. 9th International Conference on Pattern Recognition, Rome, 1988, pp. 11–16.

[2] Nicholas Ayache and Francis Lustman, "Trinocular Stereo Vision for Robotics", IEEE Transactions on PAMI, Vol 13, No. 1, January 1991, pp. 73–85.

[3] Patrick Courtney, Neil A. Thacker and Chris R. Brown, "A Hardware Architecture for Image Rectification and Ground Plane Obstacle Avoidance", Proc. 11th IAPR, ICPR 1992, pp. 23–26.

[4] 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.

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

[6] O. D. Faugeras, Q.-T. Luong and S. J. Maybank, "Camera Self-Calibration: Theory and Experiments", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992, pp. 321–334.

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

[8] 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.

[9] R. Hartley, "Cheirality Invariants", Proc. Darpa Image Understanding Workshop, April, 1993, pp. 745–753

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

[11] , "Manual of Photogrammetry, Fourth Edition", Chester C. Slama, editor. American Society of Photogrammetry, Falls Church, Va, 1980.

[12] R. Mohr, F. Veillon, L. Quan, “Relative 3D Reconstruction Using Multiple Uncalibrated Images”, Proc. CVPR-93, pp 543-548.

[13] D. V. Papadimitriou and T. J. Dennis, “Epipolar Line Estimation and Rectification for Stereo Image Pairs”, IEEE Transactions on Image Processing, Vol 5, No. 4, April 1996, pp. 672–676.

[14] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling, “Numerical Recipes in C”, Cambridge University Press, Cambridge, England, 1988

[15] Luc Robert, Michel Buffa and Martial Hébert, “Weakly-Calibrated Stereo Perception for Rover Navigation”, Proc. 5th International Conference on Computer Vision, Cambridge Mass, 1995, pp. 46 –51.

[16] V. Rodin and A. Ayache, “Axial Stereovision : Modelization and Comparison Between Two Calibration Methods”, Proc. ICIP-94, pp. 725 – 729.

[17] Fergal Shevlin, “Resampling of Scanned Imagery for Rectification and Registration”, Proc. ICIP-94, pp. 1007–1011.

[18] Rudolf Sturm, Das Problem der Projektivität und seine Anwendung auf die Flächen zweiten Grades. Math. Ann., 1:533-574, 1869.

[19] Thierry Toutin and Yves Carbonneau, “MOS and SEASAT Image Geometric Corrections”, IEEE Transactions on Geoscience and Remote Sensing, Vol 30, No. 3, May 1992, pp. 603–609

[20] George Wolberg, “Digital Image Warping”, IEEE Computer Society Press, Los Alamitos, California, 1990.

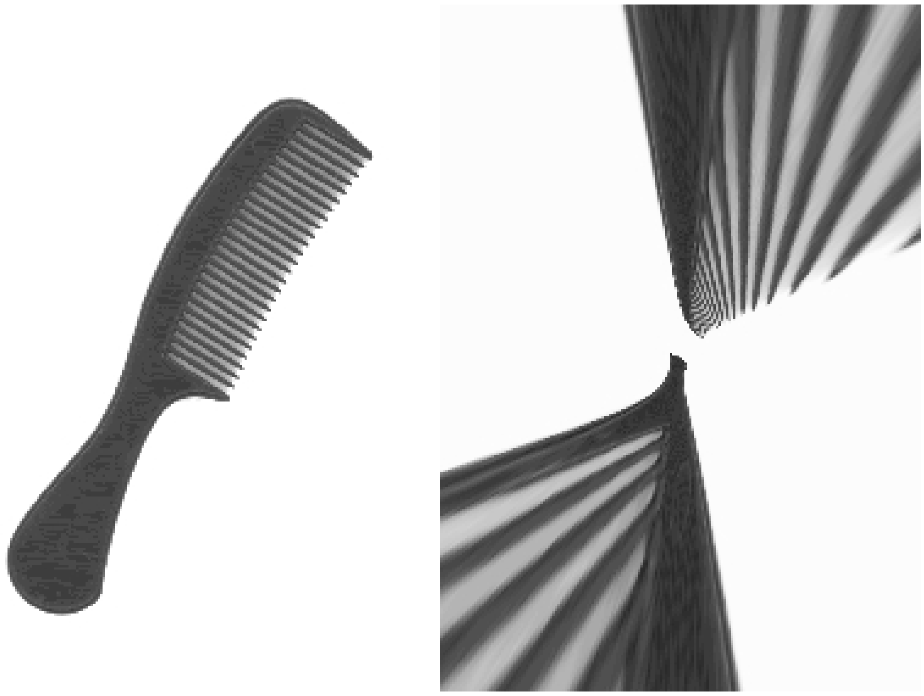

Figure 1: Picture of a comb and a non-quasi-affine resampling of the comb

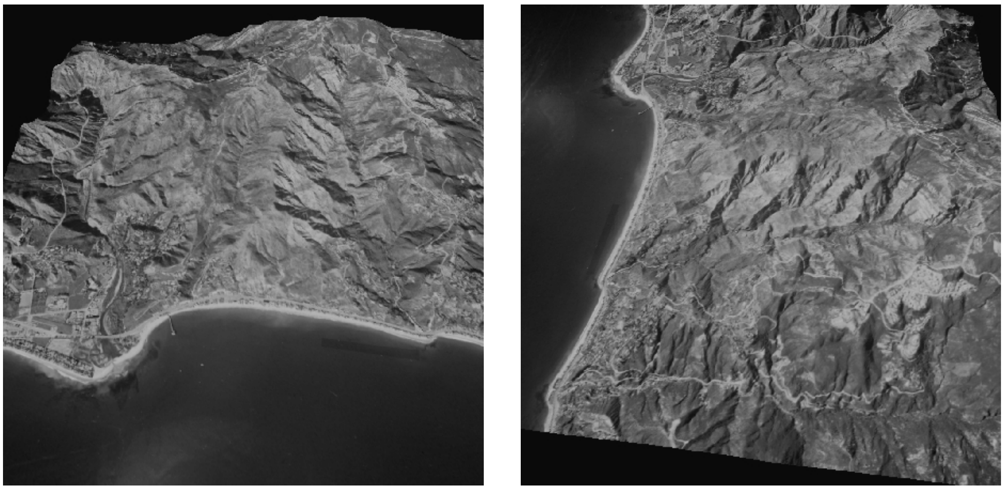

Figure 2: A pair of images of Malibu from different viewpoints.

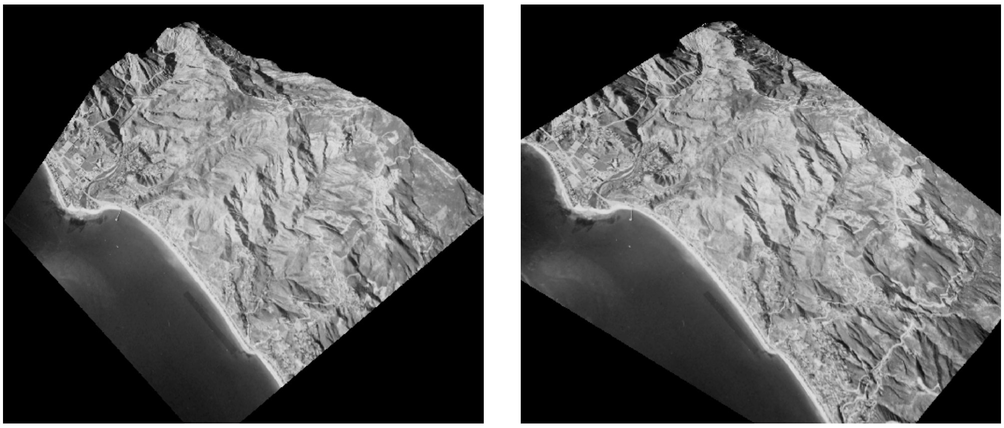

Figure 3: Pair of resampled images of Malibu

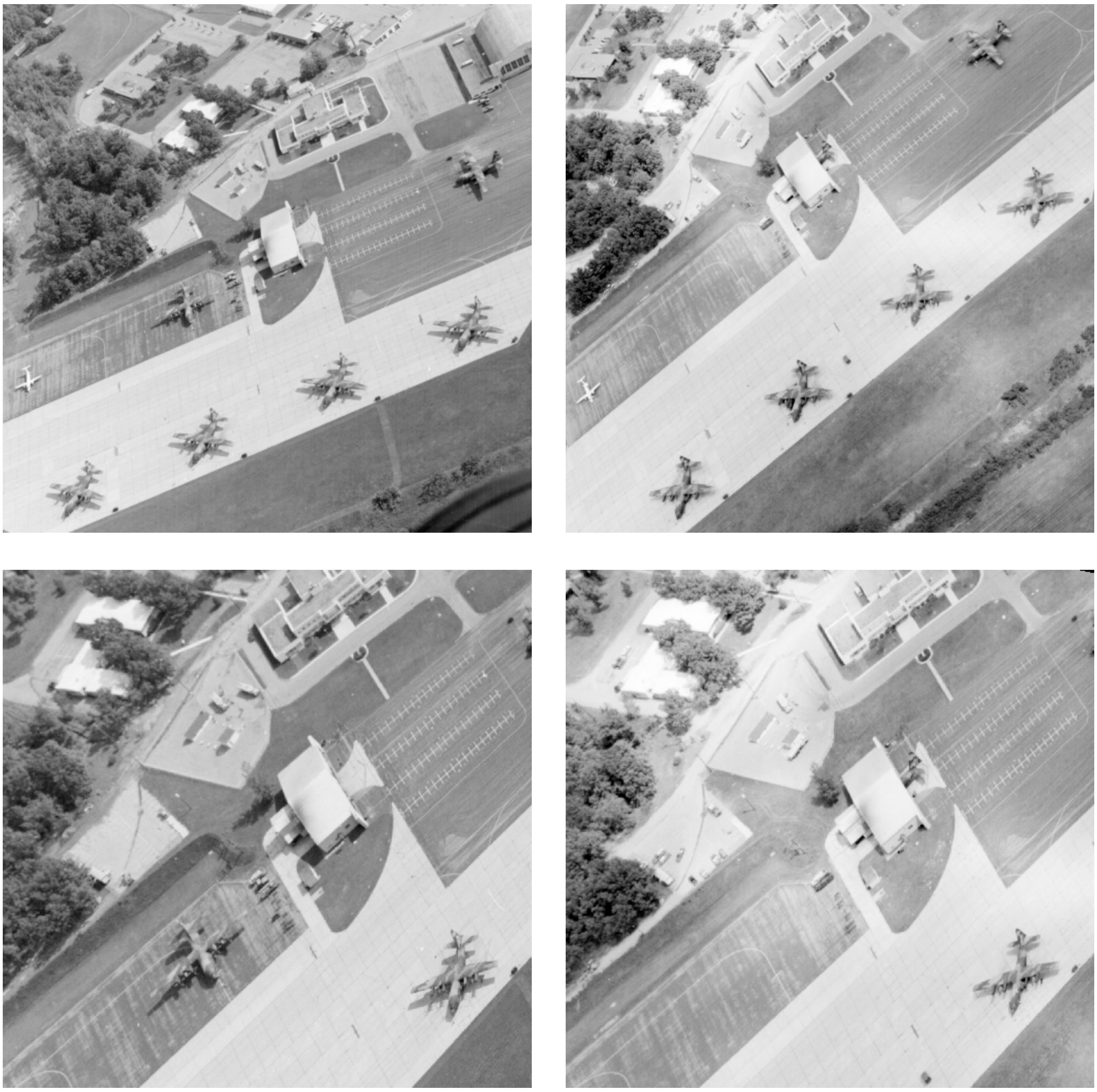

Figure 4: **Above :** A pair of images of an airfield scene taken on different days. **Below :** Rectified sections of the two images.

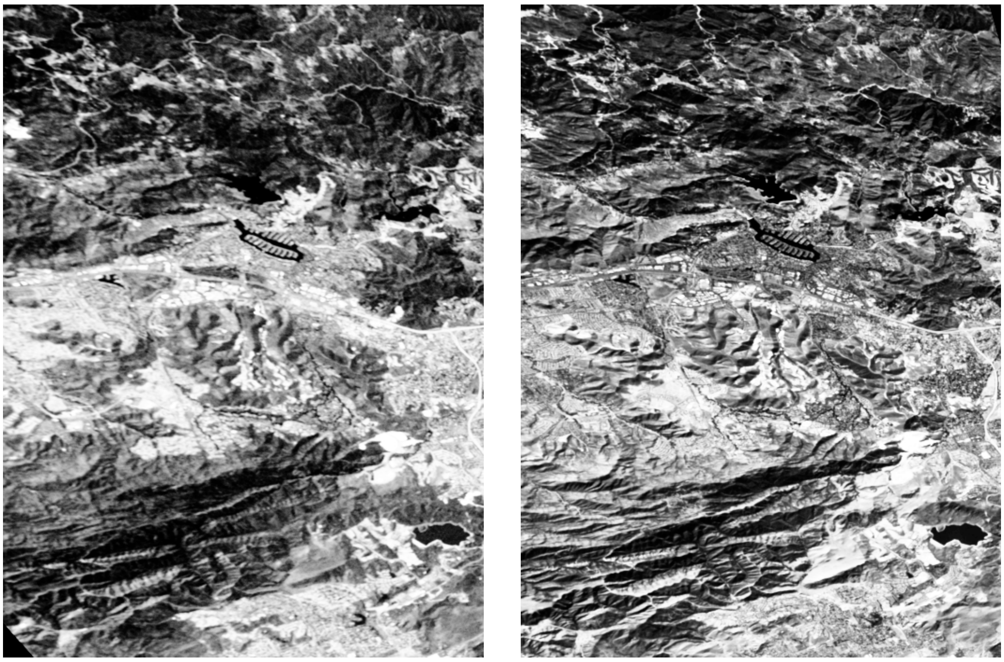

Figure 5: Resampled Landsat Image and SPOT images

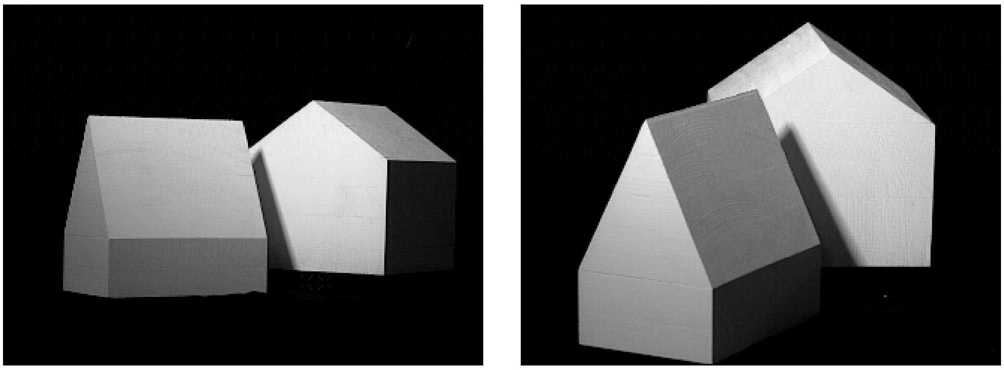

Figure 6: A pair of images of a house

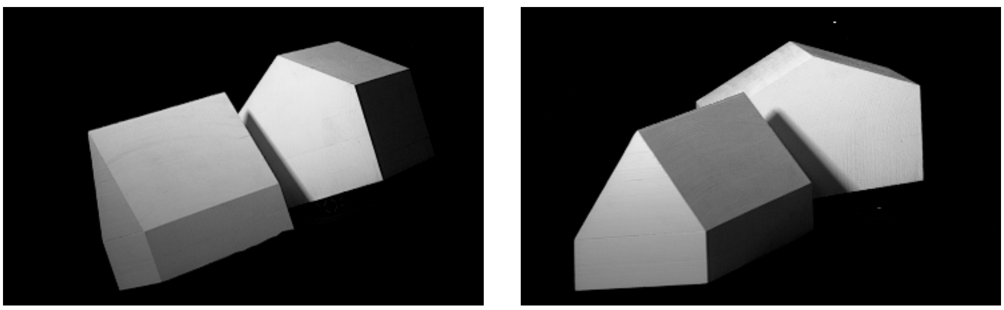

Figure 7: Resampled images of the pair of houses