Computing Rectifying Homographies for Stereo Vision

Charles Loop
Zhengyou Zhang

April 8, 1999

Technical Report
MSR-TR-99-21

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

{cloop,zhang}@microsoft.com

This paper appeared in the *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol.1, pages 125–131, June 23-25, 1999. Fort Collins, Colorado, USA.
Errors corrected on June 6, 2001.

Contents

List of Figures

Computing Rectifying Homographies for Stereo Vision

Charles Loop and Zhengyou Zhang
Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA
{cloop,zhang}@microsoft.com

June 6, 2001

Abstract

Image rectification is the process of applying a pair of 2 dimensional projective transforms, or homographies, to a pair of images whose epipolar geometry is known so that epipolar lines in the original images map to horizontally aligned lines in the transformed images. We propose a novel technique for image rectification based on geometrically well defined criteria such that image distortion due to rectification is minimized. This is achieved by decomposing each homography into a specialized projective transform, a similarity transform, followed by a shearing transform. The effect of image distortion at each stage is carefully considered.

1 Introduction

Image rectification is an important component of stereo computer vision algorithms. We assume that a pair of 2D images of a 3D object or environment are taken from two distinct viewpoints and their epipolar geometry has been determined. Corresponding points between the two images must satisfy the so-called epipolar constraint. For a given point in one image, we have to search for its correspondence in the other image along an epipolar line. In general, epipolar lines are not aligned with coordinate axis and are not parallel. Such searches are time consuming since we must compare pixels on skew lines in image space. These types of algorithms can be simplified and made more efficient if epipolar lines are axis aligned and parallel. This can be realized by applying 2D projective transforms, or *homographies*, to each image. This process is known as *image rectification*.

The pixels corresponding to point features from a rectified image pair will lie on the same horizontal scan-line and differ only in horizontal displacement. This horizontal displacement, or *disparity* between rectified feature points is related to the depth of the feature. This means that rectification can be used to recover 3D structure from an image pair without appealing to 3D geometry notions like cameras. Algorithms to find dense correspondences are based on *correlating* pixel colors along epipolar lines [1]. Seitz has shown[4] that distinct views of a scene can be *morphed* by linear interpolation along rectified scan-lines to produce new geometrically correct views of the scene.

1.1 Previous Work

Some previous techniques for finding image rectification homographies involve 3D constructions[1, 4]. These methods find the 3D line of intersection between image planes and project the two images onto the a plane containing this line that is parallel to the line joining the optical centers. While this approach is easily stated as a 3D geometric construction, its realization in practice is somewhat more involved and no consideration is given to other more *optimal* choices. A strictly 2D approach that does attempt to optimize the distorting effects of image rectification can be found in [3]. Their distortion minimization criterion is based on a simple geometric heuristic which may not lead to optimal solutions.

1.2 Overview

Our approach to rectification involves decomposing each homography into a projective and affine component. We then find the projective component that minimizes a well defined projective distortion criterion. We further decompose the affine component of each homography into a pair of simpler transforms, one designed to satisfy the constraints for rectification, the other is used to further reduce the distortion introduced by the projective component.

This paper is organized as follows. In Section 2 we present our notation and define epipolar geometry. In Section 3 we define rectification and present results needed for our homography computation. In Section 4 we give details of our decomposition. In Sections 5-7 we compute the component transforms needed for rectification. Finally, we present an example of our technique and make concluding remarks.

2 Background

We work entirely in 2 dimensional *projective space*. *Points* and *lines* are represented by lower-case bold symbols, e.g. $\mathbf{p}$ and $\mathbf{l}$. The *coordinates* of points and lines are represented by 3 dimensional column vectors, e.g. $\mathbf{p} = [\,p_u\ p_v\ p_w\,]^T$ and $\mathbf{l} = [\,l_a\ l_b\ l_c\,]^T$. The individual coordinates are sometimes ordered u, v, w for points, and a, b, c for lines. *Transforms* on points and lines are represented by 3×3 matrices associated with bold upper case symbols, e.g. $\mathbf{T}$. Unless identified to the contrary, matrix entries are given subscripted upper-case symbols, e.g. $T_{11}, T_{12}, \ldots, T_{33}$. Pure scalar quantities are given lower-case Greek symbols.

As projective quantities, points and lines are *scale invariant*, meaning $\mathbf{p} = \alpha\mathbf{p}$ ($\alpha \neq 0$) represent the same point. Points with w-coordinate equal to zero are known as *affine vectors*, *directions* or *points at* ∞. Points with a non-zero w-coordinate are known as *affine points* when the scale has been fixed so that $\mathbf{p} = [\,p_u\ \ p_v\ \ 1\,]^T$. The set of all affine points is known as the *affine plane*. For our purposes, we consider the image plane to be an affine plane where points are uniquely identified by u and v, w is presumed to be equal to one.

2.1 Epipolar Geometry

We now formally define the epipolar geometry between a pair of images. Let C and C' be a pair of *pinhole cameras* in 3D space. Let $\mathbf{m}$ and $\mathbf{m}'$ be the projections through cameras C and C' of a 3D point M in images $\mathcal{I}$ and $\mathcal{I}'$ respectively. The geometry of these definitions is shown in Fig. 1. The epipolar constraint is defined

$$\mathbf{m}'^T\mathbf{F}\mathbf{m} = 0, \tag{1}$$

for all pairs of images correspondences $\mathbf{m}$ and $\mathbf{m}'$, where $\mathbf{F}$ is the so-called *fundamental matrix* [1].

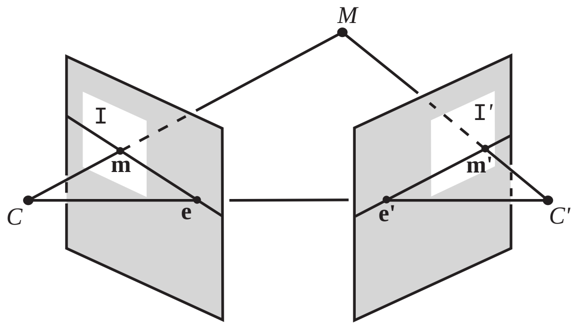


Figure 1: Epipolar geometry between a pair of images.

The fundamental matrix $\mathbf{F}$ is a 3×3 rank-2 matrix that maps points in $\mathcal{I}$ to lines in $\mathcal{I}'$, and points in $\mathcal{I}'$ to lines in $\mathcal{I}$. That is, if $\mathbf{m}$ is a point in $\mathcal{I}$ then $\mathbf{Fm} = \mathbf{l}'$ is an *epipolar line* in $\mathcal{I}'$ since from Eq. (1), $\mathbf{m}'^T\mathbf{l}' = 0$. In fact, any point $\mathbf{m}'$ that corresponds with $\mathbf{m}$ must line on the epipolar line $\mathbf{Fm}$.

For a fundamental matrix $\mathbf{F}$ there exists a pair of unique points $\mathbf{e} \in \mathcal{I}$ and $\mathbf{e}' \in \mathcal{I}'$ such that

$$\mathbf{Fe} \;=\; \mathbf{0} \;=\; \mathbf{F}^T\mathbf{e}' \tag{2}$$

where $\mathbf{0} = [\,0\ 0\ 0\,]^T$ is the *zero vector*. The points $\mathbf{e}$ and $\mathbf{e}'$ are known as the *epipoles* of images $\mathcal{I}$ and $\mathcal{I}'$ respectively. The epipoles have the property that all epipolar lines in $\mathcal{I}$ pass through $\mathbf{e}$, similarly all epipolar lines in $\mathcal{I}'$ pass through $\mathbf{e}'$.

In 3D space, $\mathbf{e}$ and $\mathbf{e}'$ are the intersections of the line $C\,C'$ with the planes containing image $\mathcal{I}$ and $\mathcal{I}'$. The set of planes containing the line $C\,C'$ are called *epipolar planes*. Any 3D point M not on line $C\,C'$ will define an epipolar plane, the intersection of this epipolar plane with the plane containing $\mathcal{I}$ or $\mathcal{I}'$ will result in an epipolar line (see Figure 1).

In this paper, we assume that $\mathbf{F}$ is known. An overview of techniques to find $\mathbf{F}$ can be found in [5]. If the intrinsic parameters of a camera are known, we say the images are calibrated, and the fundamental matrix becomes the essential matrix [1]. Our method of rectification is suitable for calibrated or uncalibrated images pairs, provided that $\mathbf{F}$ is known between them.

3 Rectification

Image rectification can be view as the process of transforming the epipolar geometry of a pair of images into a canonical form. This is accomplished by applying a homography to each image that maps the epipole to a predetermined point. We follow the convention that this point be $\mathbf{i} = [\,1\ 0\ 0\,]^T$ (a point at ∞), and that the fundamental matrix for a rectified image pair be defined

$$\bar{\mathbf{F}} = [\,\mathbf{i}\,]_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

We use the notation $[\,\mathbf{x}\,]_\times$ to denote the antisymetric matrix representing the cross product with $\mathbf{x}$. Under these conventions, it is easy to verify that rectified images have the following two properties:

i. *All epipolar lines are parallel to the u-coordinate axis,*

ii. *Corresponding points have identical v-coordinates.*

These properties are useful in practice since rectification maps epipolar lines to image scan-lines. Other conventions for canonical epipolar geometry may be useful under special circumstances.

Let $\mathbf{H}$ and $\mathbf{H}'$ be the homographies to be applied to images $\mathcal{I}$ and $\mathcal{I}'$ respectively, and let $\mathbf{m} \in \mathcal{I}$ and $\mathbf{m}' \in \mathcal{I}'$ be a pair of points that satisfy Eq. (1). Consider rectified image points $\bar{\mathbf{m}}$ and $\bar{\mathbf{m}}'$ defined

$$\bar{\mathbf{m}} = \mathbf{H}\mathbf{m} \quad \text{and} \quad \bar{\mathbf{m}}' = \mathbf{H}'\mathbf{m}'.$$

It follows from Eq. (1) that

$$\begin{aligned} \bar{\mathbf{m}}'^T \bar{\mathbf{F}} \bar{\mathbf{m}} &= 0, \\ \mathbf{m}'^T \underbrace{\mathbf{H}'^T \bar{\mathbf{F}} \mathbf{H}}_{\mathbf{F}} \mathbf{m} &= 0, \end{aligned}$$

resulting in the factorization

$$\mathbf{F} = \mathbf{H}'^T [\,\mathbf{i}\,]_\times \mathbf{H}. \tag{3}$$

Note that the homographies $\mathbf{H}$ and $\mathbf{H}'$ that satisfiy Eq. (3) are not unique. Our task is to find a pair of homographies $\mathbf{H}$ and $\mathbf{H}'$ minimize image distortion.

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be lines equated to the rows of $\mathbf{H}$ such that

$$\mathbf{H} = \begin{bmatrix} \mathbf{u}^T \\ \mathbf{v}^T \\ \mathbf{w}^T \end{bmatrix} = \begin{bmatrix} u_a & u_b & u_c \\ v_a & v_b & v_c \\ w_a & w_b & w_c \end{bmatrix}. \tag{4}$$

Similarly, let lines $\mathbf{u}'$, $\mathbf{v}'$, and $\mathbf{w}'$ be equated to the rows of $\mathbf{H}'$. By definition we have that

$$\mathbf{H}\mathbf{e} = \begin{bmatrix} \mathbf{u}^T\mathbf{e} & \mathbf{v}^T\mathbf{e} & \mathbf{w}^T\mathbf{e} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T.$$

This means that the lines $\mathbf{v}$ and $\mathbf{w}$ must contain the epipole $\mathbf{e}$. Similarly $\mathbf{v}'$ and $\mathbf{w}'$ must contain the other epipole $\mathbf{e}'$. Furthermore, we show in Appendix A that lines $\mathbf{v}$ and $\mathbf{v}'$, and lines $\mathbf{w}$ and $\mathbf{w}'$ must be corresponding epipolar lines. This has a simple geometric interpretation illustrated in Figure 2. This result establishes a linkage between the homographies $\mathbf{H}$ and $\mathbf{H}'$. This linkage is important when minimizing distortion caused by rectification.

4 Decomposition of the Homographies

We compute rectifying homographies $\mathbf{H}$ and $\mathbf{H}'$ by decomposing them into simpler transforms. Each component transform is then computed to achieve a desired effect and satisfy some conditions.

It is convenient to equate the scale invariant homography $\mathbf{H}$ with a scale variant counterpart

$$\mathbf{H} = \begin{bmatrix} u_a & u_b & u_c \\ v_a & v_b & v_c \\ w_a & w_b & 1 \end{bmatrix}, \tag{5}$$

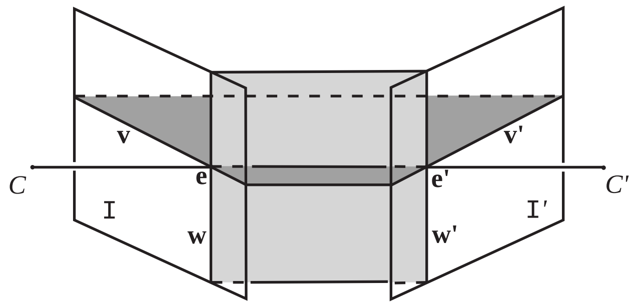


Figure 2: The lines $\mathbf{v}$ and $\mathbf{v}'$, and $\mathbf{w}$ and $\mathbf{w}'$ must be corresponding epipolar lines that lie on common epipolar planes.

by dividing out w_c. We similarly divide out w'_c from $\mathbf{H}'$. This will not lead to difficulties arising from the possibility that w_c or w'_c be equal to zero since we assume the image coordinate system origin is near the image and our minimization procedure will tend to keep the lines $\mathbf{w}$ and $\mathbf{w}'$ away from the images.

We decompose $\mathbf{H}$ into

$$\mathbf{H} = \mathbf{H}_a \mathbf{H}_p,$$

where $\mathbf{H}_p$ is a projective transform and $\mathbf{H}_a$ is an affine transform. We define

$$\mathbf{H}_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ w_a & w_b & 1 \end{bmatrix}. \tag{6}$$

From Eqs. (5) and (6) it follows that

$$\mathbf{H}_a = \mathbf{H}\mathbf{H}_p^{-1} = \begin{bmatrix} u_a - u_c w_a & u_b - u_c w_b & u_c \\ v_a - v_c w_a & v_b - v_c w_b & v_c \\ 0 & 0 & 1 \end{bmatrix}.$$

The definitions of $\mathbf{H}'_a$ and $\mathbf{H}'_p$ are similar but with primed symbols.

We further decompose $\mathbf{H}_a$ (similarly $\mathbf{H}'_a$) into

$$\mathbf{H}_a = \mathbf{H}_s \mathbf{H}_r$$

where $\mathbf{H}_r$ is similarity transformation, and $\mathbf{H}_s$ is a shearing transformation. The transform $\mathbf{H}_r$ will have the form

$$\mathbf{H}_r = \begin{bmatrix} v_b - v_c w_b & v_c w_a - v_a & 0 \\ v_a - v_c w_a & v_b - v_c w_b & v_c \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

We define $\mathbf{H}_s$ as

$$\mathbf{H}_s = \begin{bmatrix} s_a & s_b & s_c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that $\mathbf{H}_s$ only effects the u-coordinate of a point, therefore it will not effect the rectification of an image.

We now consider how to compute each component transforms just defined.

5 Projective Transform

The transforms $\mathbf{H}_p$ and $\mathbf{H}'_p$ completely characterize the projective components of $\mathbf{H}$ and $\mathbf{H}'$. These transforms map the epipoles $\mathbf{e}$ and $\mathbf{e}'$ to points at ∞ (points with w-coordinate equal to zero). By definition, $\mathbf{H}_p$ and $\mathbf{H}'_p$ are determined by lines $\mathbf{w}$ and $\mathbf{w}'$ respectively.

The lines $\mathbf{w}$ and $\mathbf{w}'$ are not independent. Given a direction $\mathbf{z} = [\,\lambda\ \mu\ 0\,]^T$ in image $\mathcal{I}$, we find

$$\mathbf{w} = [\,\mathbf{e}\,]_\times \mathbf{z} \quad \text{and} \quad \mathbf{w}' = \mathbf{F}\mathbf{z}. \tag{8}$$

This result follows from the correspondence of $\mathbf{w}$ and $\mathbf{w}'$ as epipolar lines (see Appendix A for details). Any such $\mathbf{z}$ will define a pair of corresponding epipolar lines; we are trying to find $\mathbf{z}$ that minimizes distortion, to be defined below.

Let $\mathbf{p}_i = [\,p_{i,u}\ p_{i,v}\ 1\,]^T$ be a point in the original image. This point will be transformed by $\mathbf{H}_p$ to point $[\,\frac{p_{i,u}}{w_i}\ \frac{p_{i,v}}{w_i}\ 1\,]^T$ with weight

$$w_i = \mathbf{w}^T \mathbf{p}_i.$$

If the weights assigned to points are identical then there is no projective distortion and the homography is necessarily an affine transform. In order to map the epipole $\mathbf{e}$ from the affine (image) plane to a point at ∞, $\mathbf{H}_p$ cannot in general be affine. However, as the image is bounded we can attempt to make $\mathbf{H}_p$ as affine as possible. This is the basis of our distortion minimization criterion.

5.1 Distortion Minimization Criterion

Although we cannot have identical weights in general (except when the epipole is already at ∞), we can try to minimize the variation of the weights assigned to a collection of points over both images. We use all the pixels from both images as our collection, but some other subset of important image points could also be used if necessary. The variation is measured with respect to the weight associated with the image center. More formally, we compute

$$\sum_{i-1}^{n} \left[\frac{w_i - w_c}{w_c} \right]^2, \tag{9}$$

where $w_c = \mathbf{w}^T \mathbf{p}_c$, where $\mathbf{p}_c = \frac{1}{n} \sum_{i=1}^{n} \mathbf{p}_i$ is the *average* of the points. This measure will be zero if the weights for all the points are equal, occurring only if $\mathbf{H}_p$ is an affine map, and the epipole is already at ∞. By minimizing Eq. (9) we find $\mathbf{H}_p$ and $\mathbf{H}'_p$ that are as close to affine as possible over the point set $\mathbf{p}_i$.

Over one image, Eq. (9) can be written as

$$\sum_{i=1}^{n} \left[\frac{\mathbf{w}^T (\mathbf{p}_i - \mathbf{p}_c)}{\mathbf{w}^T \mathbf{p}_c} \right]^2,$$

or as a matrix equation

$$\frac{\mathbf{w}^T \mathbf{P} \mathbf{P}^T \mathbf{w}}{\mathbf{w}^T \mathbf{p}_c \mathbf{p}_c^T \mathbf{w}}, \tag{10}$$

where $\mathbf{P}$ is the $3 \times n$ matrix

$$\mathbf{P} = \begin{bmatrix} p_{1,u} - p_{c,u} & p_{2,u} - p_{c,u} & \cdots & p_{n,u} - p_{c,u} \\ p_{1,v} - p_{c,v} & p_{2,v} - p_{c,v} & \cdots & p_{n,v} - p_{c,v} \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

We similarly define $\mathbf{p}'_c$ and $\mathbf{P}'$ for the other image.

Since $\mathbf{w} = [\,\mathbf{e}\,]_\times \mathbf{z}$ and $\mathbf{w}' = \mathbf{F}\mathbf{z}$, we rewrite Eq. (10) over both images to get

$$\frac{\mathbf{z}^T \overbrace{[\,\mathbf{e}\,]_\times^T \mathbf{P}\mathbf{P}^T [\,\mathbf{e}\,]_\times}^{\mathbf{A}} \mathbf{z}}{\mathbf{z}^T \underbrace{[\,\mathbf{e}\,]_\times^T \mathbf{p}_c \mathbf{p}_c^T [\,\mathbf{e}\,]_\times}_{\mathbf{B}} \mathbf{z}} + \frac{\mathbf{z}^T \overbrace{\mathbf{F}^T \mathbf{P}' \mathbf{P}'^T \mathbf{F}}^{\mathbf{A}'} \mathbf{z}}{\mathbf{z}^T \underbrace{\mathbf{F}^T \mathbf{p}'_c \mathbf{p}'^T_c \mathbf{F}}_{\mathbf{B}'} \mathbf{z}},$$

or simply

$$\frac{\mathbf{z}^T \mathbf{A}\mathbf{z}}{\mathbf{z}^T \mathbf{B}\mathbf{z}} + \frac{\mathbf{z}^T \mathbf{A}'\mathbf{z}}{\mathbf{z}^T \mathbf{B}'\mathbf{z}}, \tag{11}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}'$, and $\mathbf{B}'$ are 3×3 matrices that depend on the point sets $\mathbf{p}_i$ and $\mathbf{p}'_j$. Since the w-coordinate of $\mathbf{z}$ is equal to zero, only the upper-left 2×2 blocks of these matrices are important. In the remainder of this section, we denote $\mathbf{z} = [\,\lambda, \mu\,]^T$.

We now consider the specific point set corresponding to a whole image. We assume that an image is a collection of pixel locations denoted

$$\mathbf{p}_{i,j} = \begin{bmatrix} i & j & 1 \end{bmatrix}^T,$$

where $i = 0, \ldots, w-1$ and $j = 0, \ldots, h-1$, and w and h are the width and height of image $\mathcal{I}$, The image center is the point

$$\mathbf{p}_c = \begin{bmatrix} \frac{w-1}{2} & \frac{h-1}{2} & 1 \end{bmatrix}^T.$$

We similarly define primed counter-parts for image $\mathcal{I}'$. Under these assumptions, $\mathbf{P}\mathbf{P}^T$ is reduced to the following simple form:

$$\mathbf{P}\mathbf{P}^T = \frac{wh}{12} \begin{bmatrix} w^2 - 1 & 0 & 0 \\ 0 & h^2 - 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{p}_c \mathbf{p}_c^T = \frac{1}{4} \begin{bmatrix} (w-1)^2 & (w-1)(h-1) & 2(w-1) \\ (w-1)(h-1) & (h-1)^2 & 2(h-1) \\ 2(w-1) & 2(h-1) & 4 \end{bmatrix}.$$

Using these results, we compute the 2×2 matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A}'$, and $\mathbf{B}'$ in Eq. (11).

5.2 Solving the minimization problem

Solving $\mathbf{z}$ by minimizing Eq. (11) is a nonlinear optimization problem. $\mathbf{z} = [\lambda\,\mu]^T$ is defined up to a scalar factor. Without loss of generality, we can set $\mu = 1$. (If μ is much smaller than λ, we can set

$\lambda = 1$, but the following discussion still holds.) Quantity (11) is minimized when the first derivative with respect to λ is equal to 0. This gives us a polynomial of degree 7, because (11) is the sum of two rational functions, each the ratio of quadratic polynomials. The root can be found iteratively starting from an initial guess.

The initial guess is obtained as follows. We first minimize $\mathbf{z}^T\mathbf{A}\mathbf{z}/\mathbf{z}^T\mathbf{B}\mathbf{z}$ and $\mathbf{z}^T\mathbf{A}'\mathbf{z}/\mathbf{z}^T\mathbf{B}'\mathbf{z}$ separately (see below), which gives us two different estimations of $\mathbf{z}$, denoted by $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$. Their average, $(\hat{\mathbf{z}}_1/\|\hat{\mathbf{z}}_1\| + \hat{\mathbf{z}}_2/\|\hat{\mathbf{z}}_2\|)/2$, is used as the initial guess of $\mathbf{z}$. It turns out that this is very close to the optimal solution.

Minimizing $\mathbf{z}^T\mathbf{A}\mathbf{z}/\mathbf{z}^T\mathbf{B}\mathbf{z}$ is equivalent to maximizing $\mathbf{z}^T\mathbf{B}\mathbf{z}/\mathbf{z}^T\mathbf{A}\mathbf{z}$, denoted by $f(\mathbf{z})$. As $\mathbf{A}$ is symmetric and positive-definite, it can be decomposed as $\mathbf{A} = \mathbf{D}^T\mathbf{D}$. Let $\mathbf{y} = \mathbf{D}\mathbf{z}$. Then, $f(\mathbf{z})$ becomes

$$\hat{f}(\mathbf{y}) = \frac{\mathbf{y}^T\mathbf{D}^{-T}\mathbf{B}\mathbf{D}^{-1}\mathbf{y}}{\mathbf{y}^T\mathbf{y}}.$$

Since $\mathbf{y}$ is defined up to a scale factor, we can impose $\|\mathbf{y}\| = 1$, and $\hat{f}(\mathbf{y})$ is maximized when $\mathbf{y}$ is equal to the eigenvector of $\mathbf{D}^{-T}\mathbf{B}\mathbf{D}^{-1}$ associated with the largest eigenvalue. Finally, the solution for $\mathbf{z}$ is given by $\mathbf{z} = \mathbf{D}^{-1}\mathbf{y}$. Exactly the same procedure can be applied to find $\mathbf{z}$ which minimizes $\mathbf{z}^T\mathbf{A}'\mathbf{z}/\mathbf{z}^T\mathbf{B}'\mathbf{z}$.

6 Similarity Transform

In the previous section, the transforms $\mathbf{H}_p$ and $\mathbf{H}'_p$ were found that map the epipoles $\mathbf{e}$ and $\mathbf{e}'$ to points at ∞. In this section we define a pair of similarity transforms $\mathbf{H}_r$ and $\mathbf{H}'_r$ that *rotate* these points at ∞ into alignment with the direction $\mathbf{i} = [1\ 0\ 0]^T$ as required for rectification. Additionally, a translation in the v-direction on one of the images is found to exactly align the scan-lines in both images.

At this stage, we assume that the lines $\mathbf{w}$ and $\mathbf{w}'$ are known. We can therefore eliminate v_a and v_b from Eq. (7) by making use of the following:

$$\begin{aligned}\mathbf{F} &= \mathbf{H}'^T[\mathbf{i}]_\times\mathbf{H} \qquad (12)\\ &= \begin{bmatrix} v_a w'_a - v'_a w_a & v_b w'_a - v'_a w_b & v_c w'_a - v'_a \\ v_a w'_b - v'_b w_a & v_b w'_b - v'_b w_b & v_c w'_b - v'_b \\ v_a - v'_c w_a & v_b - v'_c w_b & v_c - v'_c \end{bmatrix}.\end{aligned}$$

Using the last row of this matrix equation, we determine that

$$\begin{aligned} v_a &= F_{31} + v'_c w_a, \qquad (13)\\ v_b &= F_{32} + v'_c w_b, \qquad (14)\\ v_c &= F_{33} + v'_c. \qquad (15)\end{aligned}$$

Eqs. (13-15) are substituted into Eq (7), resulting in

$$\mathbf{H}_r = \begin{bmatrix} F_{32} - w_b F_{33} & w_a F_{33} - F_{31} & 0 \\ F_{31} - w_a F_{33} & F_{32} - w_b F_{33} & F_{33} + v'_c \\ 0 & 0 & 1 \end{bmatrix}. \qquad (16)$$

Similarly, v'_a and v'_b can be eliminated to get

$$\mathbf{H}'_r = \begin{bmatrix} w'_b F_{33} - F_{23} & F_{13} - w'_a F_{33} & 0 \\ w'_a F_{33} - F_{13} & w'_b F_{33} - F_{23} & v'_c \\ 0 & 0 & 1 \end{bmatrix}. \tag{17}$$

Note that there remains a translation term involving v'_c in Eqs. (17) and (16). This shows that translation in the v-direction is linked between the two images, and that an offset of F_{33} is needed to align horizontal scan-lines. We find v'_c so that the minimum v-coordinate of a pixel in either image is zero.

As similarity transforms, $\mathbf{H}_r$ and $\mathbf{H}'_r$ can only rotate, translate, and uniformly scale the images $\mathcal{I}$ and $\mathcal{I}'$. None of these operations introduce any additional distortion.

The combined transforms $\mathbf{H}_r\mathbf{H}_p$, and $\mathbf{H}'_r\mathbf{H}'_p$ are sufficient to rectify images $\mathcal{I}$ and $\mathcal{I}'$. However, there remains additional freedom, corresponding to $\mathbf{u}$ and $\mathbf{u}'$ of Eq. (4). These elements take the form of shearing transforms described below, that can be leveraged to reduce distortion and to map the images into a more practical pixel range.

7 Shearing Transform

In this section the freedom afforded by the independence of $\mathbf{u}$ and $\mathbf{u}'$ is exploited to reduce the distortion introduced by the projective transforms $\mathbf{H}_p$ and $\mathbf{H}'_p$. Due to this independence, we consider only one image, as the procedure is carried out identically on both images.

We model the effect of $\mathbf{u}$ as a shearing transform

$$\mathbf{S} = \begin{bmatrix} a & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We set the translation components of $\mathbf{S}$ to zero since these terms add no useful freedom at this stage.

Let $\mathbf{a} = [\, \frac{w-1}{2} \;\; 0 \;\; 1 \,]^T$, $\mathbf{b} = [\, w-1 \;\; \frac{h-1}{2} \;\; 1 \,]^T$, $\mathbf{c} = [\, \frac{w-1}{2} \;\; h-1 \;\; 1 \,]^T$, and $\mathbf{d} = [\, 0 \;\; \frac{h-1}{2} \;\; 1 \,]^T$ be points corresponding to the midpoints of the edges of $\mathcal{I}$. Furthermore, let $\hat{\mathbf{a}} = \mathbf{H}_r\mathbf{H}_p\mathbf{a}$, be a point in the affine plane by dividing through so that $\hat{\mathbf{a}}_w = 1$; similarly define $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, and $\hat{\mathbf{d}}$.

In general, $\mathbf{H}_p$ is a projective transform, so it is not possible to undistort $\mathcal{I}$ completely using the affine transform $\mathbf{S}$. Instead we attempt to preserve perpendicularity and aspect ratio of the lines $\bar{\mathbf{bd}}$ and $\bar{\mathbf{ca}}$. Let

$$\begin{aligned} \mathbf{x} &= \hat{\mathbf{b}} - \hat{\mathbf{d}}, \\ \mathbf{y} &= \hat{\mathbf{c}} - \hat{\mathbf{a}}. \end{aligned}$$

As the difference of affine points, $\mathbf{x}$ and $\mathbf{y}$ are vectors in the euclidean image plane. Perpendicularity is preserved when

$$(\mathbf{Sx})^T(\mathbf{Sy}) = 0, \tag{18}$$

and aspect ratio is preserved if

$$\frac{(\mathbf{Sx})^T(\mathbf{Sx})}{(\mathbf{Sy})^T(\mathbf{Sy})} = \frac{w^2}{h^2}. \tag{19}$$

Eqs. (18) and (19) represent quadratic polynomials in a and b (the unknown elements of $\mathbf{S}$) whose simultaneous satisfaction is required. Using the method outlined in [2] we find the real solution

$$a = \frac{h^2 x_v^2 + w^2 y_v^2}{hw(x_v y_u - x_u y_v)} \quad \text{and} \quad b = \frac{h^2 x_u x_v + w^2 y_u y_v}{hw(x_u y_v - x_v y_u)},$$

up to sign; the solution where a is positive is preferred. We define $\mathbf{H}_s$ (and similarly $\mathbf{H}'_s$) to be $\mathbf{S}$ composed with a uniform scaling and translation as discribed below.

The combined transform $\mathbf{H}_s\mathbf{H}_r\mathbf{H}_p$, and similarly $\mathbf{H}'_s\mathbf{H}'_r\mathbf{H}'_p$, rectify images $\mathcal{I}$ and $\mathcal{I}'$ with minimal distortion. However these image may not the appropriate size, or in the most desirable coordinate system. Therefore, additional uniform scaling and translation may be applied. It is important that the same scale factor, and the same v-translation be applied to both images to preserve rectification. Translations in the u-direction have no effect on rectification.

In our examples, we chose a uniform scale factor that preserves the sum of image areas. Other criteria may work equally well. We also compute translations in u so that the minimum pixel coordinate has a u-coordinate of zero. A similar translation is found for the v direction, but the minimum is taken over both images to preserve rectification.

8 Conclusion

We have presented a procedure for computing rectification homographies for a pair of images taken from distinct viewpoints of a 3D scene. Figure 3 shows the results of each stage of our technique on one example. This new method is based entirely on quantifiable 2D image measures and requires no 3D constructions. Furthermore, these measure have intuitive geometric meaning. We have shown the technique that minimizes distortion due to the projective component of rectification, and used additional degrees of freedom in the affine component to further reduce distortion to a well defined minimum.

A Proof of Correspondence Properties

In this appendix we demonstrate i) how corresponding epipolar lines are related by a direction in one image, and ii) that the second and third rows of a pair of rectifying homographies correspond to a pairs of corresponding epipolar lines. We use the symbol $\sim$ to indicate correspondence.

Let $\mathbf{l} \in \mathcal{I}$ and $\mathbf{l}' \in \mathcal{I}'$ be a pair of epipolar lines.

Proposition 1. *If* $\mathbf{l} \sim \mathbf{l}'$ *and* $\mathbf{x} \in \mathcal{I}$ *is a direction (point at* ∞*) such that* $\mathbf{l} = [\mathbf{e}]_\times \mathbf{x}$ *then*

$$\mathbf{l}' = \mathbf{F}\mathbf{x}.$$

Proof. Let $\mathbf{x}$ be the intersection of lines $\mathbf{l}$ and $\mathbf{k} = [\,0\ \ 0\ \ 1\,]^T$ (the line at ∞), found by $\mathbf{x} = [\mathbf{k}]_\times \mathbf{l}$. Similarly, let $\mathbf{x}' = [\mathbf{k}]_\times \mathbf{l}'$. Clearly $\mathbf{l} = [\mathbf{e}]_\times \mathbf{x}$, since $[\,\mathbf{e}\,]_\times [\,\mathbf{k}\,]_\times \mathbf{l} = \mathbf{l}$.

Since $\mathbf{l} \sim \mathbf{l}'$ it follows that $\mathbf{x}'^T \mathbf{F}\mathbf{x} = 0$. By defintion $\mathbf{e}'^T \mathbf{l}' = \mathbf{e}'^T \mathbf{F}\mathbf{x} = 0$ and $\mathbf{x}'^T \mathbf{l}' = \mathbf{x}'^T \mathbf{F}\mathbf{x} = 0$, which shows that lines $\mathbf{l}'$ and $\mathbf{F}\mathbf{x}$ both contain points $\mathbf{e}'$ and $\mathbf{x}'$ and must be the same line. $\square$

In the following, we denote the rows of $\mathbf{H}$ and $\mathbf{H}'$ as in Eq. (4), and $\mathbf{i} = [\,1\ 0\ 0\,]^T$.

Proposition 2. *If* $\mathbf{H}$ *and* $\mathbf{H}'$ *are homgraphies such that*

$$\mathbf{F} = \mathbf{H}'^T [\mathbf{i}]_\times \mathbf{H}, \tag{20}$$

then $\mathbf{v} \sim \mathbf{v}'$ *and* $\mathbf{w} \sim \mathbf{w}'$.

Proof. Expanding $\mathbf{H}'^T [\mathbf{i}]_\times \mathbf{H}$ shows that

$$\mathbf{F} = \begin{bmatrix} v_a w'_a - v'_a w_a & v_b w'_a - v'_a w_b & v_c w'_a - v'_a w_c \\ v_a w'_b - v'_b w_a & v_b w'_b - v'_b w_b & v_c w'_b - v'_b w_c \\ v_a w'_c - v'_c w_a & v_b w'_c - v'_c w_b & v_c w'_c - v'_c w_c \end{bmatrix}.$$

We observe that $\mathbf{F}$ does not depend on $\mathbf{u}$ or $\mathbf{u}'$. Without loss of generality, we set $\mathbf{u} = \mathbf{k} = \mathbf{u}'$, where $\mathbf{k} = [\,0\ 0\ 1\,]^T$ is the line at ∞. It is straightforward to show that, up to a scale factor

$$\mathbf{H}^{-1} = \left[\ [\,\mathbf{v}\,]_\times \mathbf{w} \quad [\,\mathbf{w}\,]_\times \mathbf{u} \quad [\,\mathbf{u}\,]_\times \mathbf{v}\ \right].$$

Since $\mathbf{v}$ and $\mathbf{w}$ are independent (follows from Eq. (20)) and both contain $\mathbf{e}$, we conclude that $[\,\mathbf{v}\,]_\times \mathbf{w} = \mathbf{e}$. Let $\mathbf{y} = [\,\mathbf{v}\,]_\times \mathbf{k}$ and $\mathbf{z} = [\,\mathbf{w}\,]_\times \mathbf{k}$. From Eq. (20) we get

$$\begin{aligned} \mathbf{H}'^T [\mathbf{i}]_\times &= \mathbf{F}\mathbf{H}^{-1} \\ \left[\ \mathbf{k} \quad \mathbf{v}' \quad \mathbf{w}'\ \right] [\mathbf{i}]_\times &= \mathbf{F} \left[\ \mathbf{e} \quad \mathbf{z} \quad -\mathbf{y}\ \right] \\ \left[\ \mathbf{0} \quad \mathbf{w}' \quad -\mathbf{v}'\ \right] &= \left[\ \mathbf{0} \quad \mathbf{F}\mathbf{z} \quad -\mathbf{F}\mathbf{y}\ \right]. \end{aligned} \tag{21}$$

We conclude that $\mathbf{v}' = \mathbf{F}\mathbf{y}$ and $\mathbf{w}' = \mathbf{F}\mathbf{z}$. By Proposition 1 it follows that $\mathbf{v} \sim \mathbf{v}'$ and $\mathbf{w} \sim \mathbf{w}'$. $\square$

References

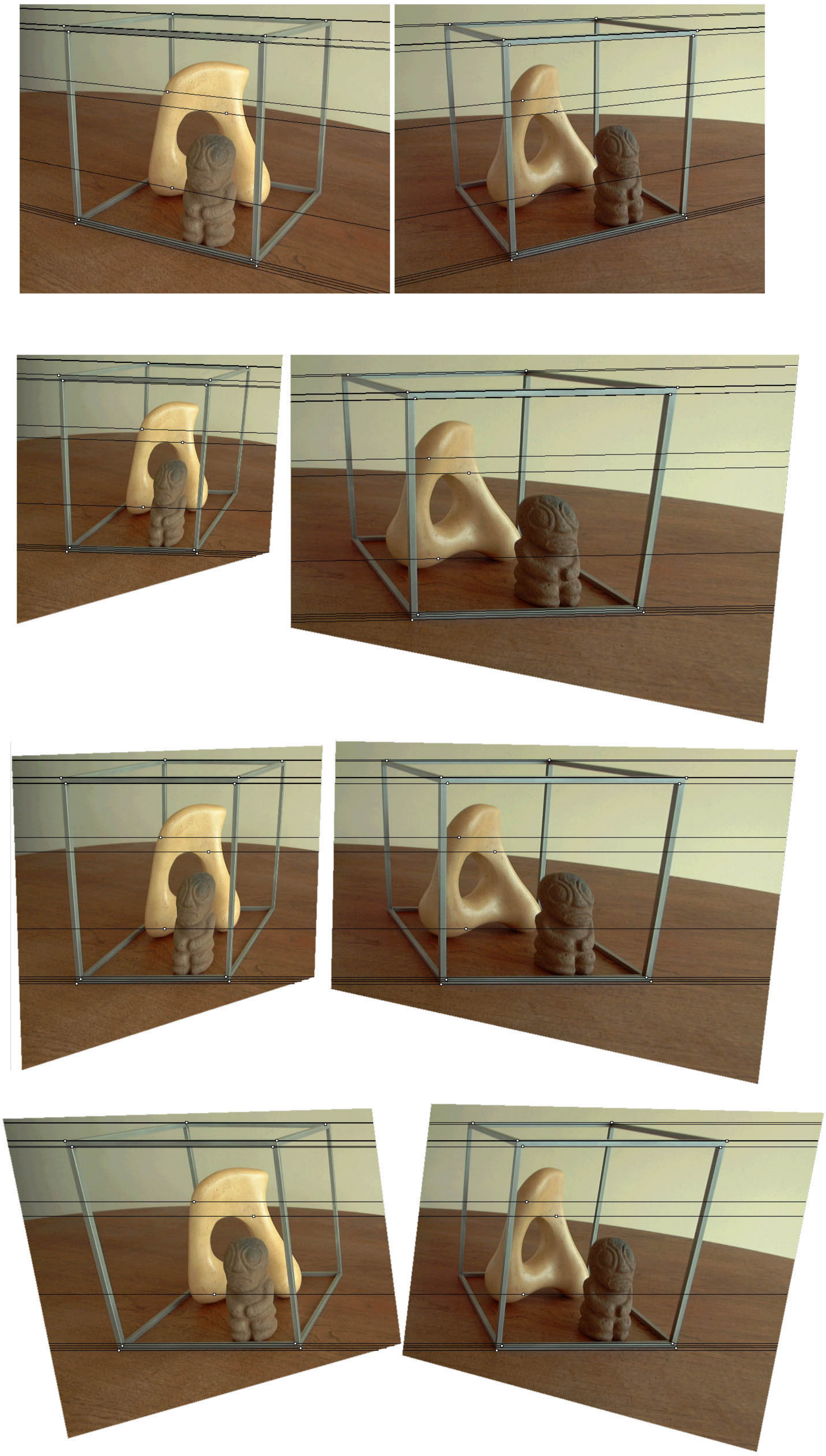
[1] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[2] D. Manocha and J Demmel. Algorithms for intersecting parametric and algebraic curves I: Simple intersections. *ACM Transactions on Graphics*, 13(1):73–100, 1994.

[3] L. Robert, C. Zeller, O. Faugeras, and M. Hébert. Applications of non-metric vision to some visually-guided robotics tasks. In Y. Aloimonos, editor, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 5, pages 89–134. Lawrence Erlbaum Associates, 1997.

[4] S. Seitz. *Image-Based Transformation of Viewpoint and Scene Appearance*. PhD thesis, University of Wisconsin, 1997.

[5] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–1195, 1998.



(a) Original image pair overlayed with several epipolar lines.

(b) Image pair transformed by the specialized projective mapping $\mathbf{H}_p$ and $\mathbf{H}'_p$. Note that the epipolar lines are now parallel to each other in each image.

(c) Image pair transformed by the similarity $\mathbf{H}_r$ and $\mathbf{H}'_r$. Note that the image pair is now rectified (the epipolar lines are horizontally aligned).

(d) Final image rectification after shearing transform $\mathbf{H}_s$ and $\mathbf{H}'_s$. Note that the image pair remains rectified, but the horizontal distortion is reduced.

Figure 3: An example showing various stages of the proposed rectification algorithm.