Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 27 (2006) 1447–1456
Pattern Recognition Letters
www.elsevier.com/locate/patrec

A method of visual metrology from uncalibrated images

Zezhi Chen *, Nick Pears, Bojian Liang

Department of Computer Science, University of York, York YO10 5DD, UK

Received 26 May 2005; received in revised form 28 October 2005
Available online 30 May 2006

Communicated by Prof. H.H.S. Ip

Abstract

A method of measuring the height of any feature above a reference plane from a pair of uncalibrated images, separated by a (near) pure translation is presented. The output of the algorithm is a feature height, expressed as a fraction of the height of the camera above the reference plane. There are three contributions. Firstly a robust method of computing the dual epipole or focus of expansion (FOE) under pure translation is presented. Secondly, a novel reciprocal-polar (RP) image rectification scheme is presented, which allows planar image motion, expressed as a planar homography, to be accurately detected and recovered by 1D correlation. The technique can work even when there are no corner features on the reference plane and even over large image distortions caused by large camera motion, which would cause correlation techniques in the original image space to fail. Thirdly, we present a projective construct to enable measurement of the relative (or affine) feature height. Results show that our algorithm performs very well against outliers and noise. The mean of absolute error is 1.8 mm, and the mean of relative error is only 0.13% with two outliers removed.

Keywords: Reciprocal-polar (RP) transform; Homography matrix; Fundamental matrix; Metrology; Focus of expansion (FOE); Uncalibrated image

1. Introduction

The goal of this paper is to be able to measure the height of features above a reference plane from two uncalibrated images separated by a (near) pure translation. We have used this to detect obstacles in a mobile robot application and measure the height of doorways, windows and furniture in our laboratory in a visual metrology application. In both of these applications we use the laboratory floor as our reference plane. In the mobile robot application, we know the height of the camera above the reference plane, whereas in the metrology experiments, we know the height of one object in the image. There are several different methods of visual metrology in the literature. For example, 3D world structure computed from uncalibrated views of a scene, given sufficient correspondences in general position, has already been in use for answering specific metric questions about the scene. The approach used by Tomasi and Kanade (1992) is known as the factorization method; Triggs (1996) later extended the factorization method to a projective camera model using epipolar constraints to calculate depth scale factors; Heyden and Berthilsson (1999) upgraded the affine approximation method to projective results employing iterative optimisation techniques. Another method is the camera-centered approach (Pollefeys et al., 1998) in which the first image is used as a reference to determine the projection matrices of other images in a projective frame under multiple geometric constraints. The world-centered approach is suitable for long image sequences while the camera-centered approach on the other hand is suitable for short image sequences. The photogrammetry technique concentrates on accuracy problems, whereas its derived techniques produces highly accurate three-dimensional models (Atkinson, 1996). However, photogrammetry techniques generally require heavy human interaction.

* Corresponding author. Tel.: +44 (0)131 451 4179; fax: +44 (0)131 451 3327.
E-mail addresses: chen@macs.hw.ac.uk (Z. Chen), nep@cs.york.ac.uk (N. Pears), bojian@cs.york.ac.uk (B. Liang).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.01.017

Many researchers have produced a number of techniques for measuring objects from images. Criminisi et al. (1999, 2000) proposed several methods to make measurement based on world planes from their perspective images. Reid and Zisserman (1996) and Kim et al. (1998) gave some methods for locating the 3D position and calculating the height of a soccer ball from monocular image sequences of soccer games. Liang and Pears (2002a,b) presented a method for computing the height of obstacles for robot visual navigation.

It is well known that, for a plane in general position, the homography is determined uniquely by the plane and vice versa. A homography can be determined by four correspondences in general position (with no three collinear correspondences) and is described by a 3×3 matrix $\boldsymbol{H}$. Under pure camera translation in some direction $\boldsymbol{t}$, image motion sources or sinks to a point $\boldsymbol{v}$ in the direction of $\boldsymbol{t}$, which is termed the focus of expansion (FOE). It is evident that $\boldsymbol{v}$ is also the epipole $\boldsymbol{e}$ and $\boldsymbol{e}'$ for both views and the epipolar lines are radial and centred on the FOE. In this situation, it is possible to carry out an affine reconstruction using two images. The geometric significance is that the vanishing line of the reference plane can be determined by two pairs of correspondences, and the normal of ground plane can be determined by orientation the vanishing line (Hartley and Zisserman, 2001; Faugeras et al., 2001). The recovered H-matrix, in conjunction with a cross-ratio construct, is further applied to measure the height of a visual feature to the reference plane. The measurement method developed here is independent of the camera's intrinsic parameters (focal length, aspect ratio, principle point, skew) and extrinsic parameter (translation $\boldsymbol{t}$). The method does not require any geometric constraints on the 3D scene and the only limiting aspect of the method is that it is only applicable to (near) pure translation. However, in many applications, such as mobile robot navigation, deliberative (near) translation motion can be executed to probe the environment in terms of height, and measured heights of near zero are obstacle-free, navigable zones.

Six salient aspects of this method are as follows: (i) A robust method for estimating the FOE was developed yielding a highly accurate result. (ii) A new reciprocal-polar (RP) image rectification renders all planar motion to a pure shift, when the camera translation direction is parallel to the plane. This allows correlation based planar image motion recovery even over large perspective image distortions induced by large camera motions. (iii) Since the image motion recovery is correlation based, no corner feature correspondences are required on the reference plane, just local intensity variation. All coplanar pixels with local intensity variation contribute to the reference plane homography estimation, not just a potentially low density of corner feature matches. (iv) We show that the magnitude of image motion in the $1/r$ dimension of the rectified image pair follows a sinusoidal form along the θ dimension over a maximum of π radians, for the four degrees-of-freedom (dof) class of planar homographies called elations. (v) Random sample consensus (RANSAC) and least squares (LS) estimation of the phase of the sinusoid yields a highly accurate vanishing line orientation of the reference plane (and simultaneously a segmentation of the reference plane) and this, along with the FOE, allows an accurate reference plane homography relation to be obtained. (vi) Finally, we present a new projective construction to compute affine height via the plane-and-parallax cue.

The paper is structured as follows. In the following section, we describe the relation between the fundamental matrix and the homography under pure translation. In Section 3, we make an incremental contribution to the estimation of the FOE or epipoles, by suggesting that a linear estimate should be refined using an optimisation based on the fundamental matrix. Section 4 presents the main result of the paper, where the analysis suggests the use of a reciprocal-polar rectification to recover the homography (H-matrix) representing reference plane motion across the image pair. Section 5 gives the result for computing affine height via the plane-and-parallax cue. Section 6 validates the method through several experiments, first using simple point correspondences and then using 1D intensity correlation, which is where the power of the method lies. Also, it is shown that, in terms of accuracy of the height, the method performs as well as the state-of-the-art. In the final section, conclusions are drawn.

2. The relation of the F and H matrix under pure translation

Images of points on a plane are related to corresponding image points in the second view by a planar homography (Hartley and Zisserman, 2001; Faugeras et al., 2001; Forsyth and Ponce, 2003). This is a projective relation since it depends only on the intersections of planes with lines. The homography map transfers points from one view to the other as if they were images of points on the plane. There are then two relations between the two views: first, through the epipolar geometry a point in one view determines a line in the other which is the image of the ray through that point; and second, through the homography a point in one view determines a point in the other which is the image of the intersection of the ray with a plane.

In pure translation of the camera, Eq. (1) gives the relation of the fundamental matrix, $\boldsymbol{F}$, and the epipoles $\boldsymbol{e}$ and $\boldsymbol{e}'$.

$$\boldsymbol{F} = [\boldsymbol{e}']_\times = [\boldsymbol{e}]_\times = [\boldsymbol{v}]_\times \tag{1}$$

where $[\cdot]_\times$ is a skew-symmetric matrix corresponding to the vector and prime denotes the second view in the image pair.

Let $\boldsymbol{X}_i$ be a set of points, which are coplanar in 3D Euclidean space. The images of $\boldsymbol{X}_i$ from two view points are related by a plane to plane projectivity or homography $\boldsymbol{H}$, such that,

$$\lambda \boldsymbol{x}'_i = \boldsymbol{H}\boldsymbol{x}_i \tag{2}$$

where λ is a scalar, $\boldsymbol{x}_i$ and $\boldsymbol{x}'_i$ are homogenous image coordinates of the images of point $\boldsymbol{X}_i$, and $\boldsymbol{H}$ is a 3×3 matrix representing the homography. As homogenous coordinates are defined up to a scale factor, the H-matrix has only eight degrees-of-freedom (dof), and it can be determined by standard linear methods of four corresponding point pairs in general position (no three collinear). When the number of point pairs is more than four, a standard least square method or SVD (singular value decomposition) method can be used.

However, two corresponding point pairs fully determine the H-matrix under pure translation parallel to the plane. Suppose the cameras are calibrated with the origin of world coordinate system at the first camera, and the intrinsic parameters, $\boldsymbol{K}$, constant. The camera matrices, $\boldsymbol{P}$, $\boldsymbol{P}'$, for the two views are

$$\boldsymbol{P} = \boldsymbol{K}[\boldsymbol{I}|\boldsymbol{0}] \quad \boldsymbol{P}' = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{T}] \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{R}$ and $\boldsymbol{T}$ represent the rotational and translation displacement between the two views respectively. If the world plane $\boldsymbol{\pi}_E$ has normal, $\boldsymbol{n}$, and distance to origin, d, so that its coordinates are $\boldsymbol{\pi}_E = (\boldsymbol{n}^{\mathrm{T}}, d)^{\mathrm{T}}$, then

$$\boldsymbol{H} = \boldsymbol{K}(\boldsymbol{R} - \lambda \boldsymbol{t}\boldsymbol{n}^{\mathrm{T}})\boldsymbol{K}^{-1} \tag{4}$$

where $\lambda = \frac{1}{d}$, For a pure translation, $\boldsymbol{R} = \boldsymbol{I}$, and so $\boldsymbol{H}$ has the form

$$\boldsymbol{H} = \boldsymbol{I} - \lambda(\boldsymbol{K}\boldsymbol{t})(\boldsymbol{K}^{-\mathrm{T}}\boldsymbol{n})^{\mathrm{T}} \tag{5}$$

We note that $\boldsymbol{K}\boldsymbol{t}$ is the FOE $\boldsymbol{v} = \eta(x_f \quad y_f \quad 1)^{\mathrm{T}}$, and $\boldsymbol{K}^{-\mathrm{T}}\boldsymbol{n}$ is the vanishing line $\boldsymbol{l} = \upsilon(a_v \quad b_v \quad 1)^{\mathrm{T}}$ corresponding to the plane $\boldsymbol{\pi}_E$. Thus, we have

$$\boldsymbol{H} = \boldsymbol{I} - k\boldsymbol{v}\boldsymbol{l}^{\mathrm{T}} \tag{6}$$

where k is a constant scalar. Since two corresponding point pairs fully define the FOE and vanishing line, so the H-matrix can, in theory, also be fully determined by two corresponding matches (Fig. 1).

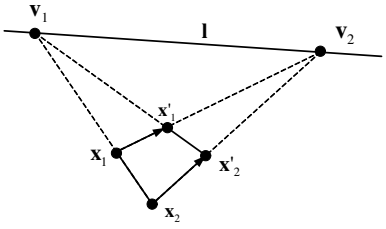


Fig. 1. Two corresponding point pairs fully define the vanishing point $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and vanishing line $\boldsymbol{l}$.

3. Accurate estimation of the FOE

We can detect (near) pure translation by intersecting all lines defined by all corner correspondences from the image pair and if great percentage of intersections lies in a small area (such as 85% of intersection should lie within a 50 pixels radius), then pure translation is assumed and the mean of the intersections is an initial value of FOE. The question now is: how can we calculate the FOE with high accuracy and high stability? A robust algorithm for estimating the FOE is summarized in Table 1.

4. Reciprocal-polar transform and recovery of H-matrix

Once the FOE has been computed, we shift image coordinates so that each image is centred on the FOE $\boldsymbol{v} = (x_f \quad y_f \quad 1)^{\mathrm{T}}$. Let

Table 1
A robust method to estimate FOE

(1) *Extract interest points*: Compute interest points in each image by using the Plessey–Harris corner detector (Harris and Stephens, 1988) or KLT algorithm (Shi and Tomasi, 1994) or SUSAN (Smith and Brady, 1997) method
(2) *Putative correspondences*: Compute a set of interest point matches based on proximity and similarity of their intensity neighbourhood
(3) *RANSAC robust estimation*: Repeat for m samples, where m is determined adaptively by using binning technique (Zhang, 1998)
  (a) Select a random sample of at least 2 correspondences and compute the FOE by using the simultaneous equations:

$$\begin{cases} (\boldsymbol{x}_1 \times \boldsymbol{x}'_1) \cdot \boldsymbol{v} = 0 \\ (\boldsymbol{x}_2 \times \boldsymbol{x}'_2) \cdot \boldsymbol{v} = 0 \\ \quad \vdots \\ (\boldsymbol{x}_n \times \boldsymbol{x}'_n) \cdot \boldsymbol{v} = 0 \end{cases}$$

  where $\boldsymbol{x}_i \leftrightarrow \boldsymbol{x}'_i$ $(i = 1, 2, \cdots, n)$ is any pair of matching points in two images
  (b) Calculate the epipolar distance $f(\boldsymbol{v})$ for each putative correspondence

$$f(\boldsymbol{v}) = \left(\frac{1}{\sqrt{(\boldsymbol{F}\boldsymbol{x}_i)_1^2 + (\boldsymbol{F}\boldsymbol{x}_i)_2^2}} + \frac{1}{\sqrt{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{x}'_i)_1^2 + (\boldsymbol{F}^{\mathrm{T}}\boldsymbol{x}'_i)_2^2}} \right) |\boldsymbol{x}_i'^{\mathrm{T}}\boldsymbol{F}\boldsymbol{x}_i|$$

  where $\boldsymbol{F}$ is the fundamental matrix and the subscripts 1, 2 denote vector components.
  (c) Compute the number of *inliers* consistent with $\boldsymbol{v}$ by the number of correspondences for which $f(\boldsymbol{v}) <$ *threshold*. Choose the FOE with the largest number of *inliers*
(4) *Optimal estimation*: Re-estimate the FOE from all correspondences classified as *inliers*, by minimizing the object function $f(\boldsymbol{v})$
(5) *Repeat* steps (3)–(4) until the number of correspondences are stable

$$\boldsymbol{T}_c = \begin{bmatrix} 1 & 0 & -x_f \\ 0 & 1 & -y_f \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

After translation $\boldsymbol{T}_c$ is applied, the FOE is at homogenous coordinates $\boldsymbol{v}_c = (0,0,1)^T$ and vanishing line becomes $\boldsymbol{l}_c = \boldsymbol{T}_c^{-T}\boldsymbol{l} = (a_v, b_v, \boldsymbol{v}^T\boldsymbol{l})^T$. The homography relating points in FOE centred coordinates is

$$\boldsymbol{H}_c = \boldsymbol{I} - k\boldsymbol{v}_c\boldsymbol{l}_c^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -ka_v & -kb_v & (1 - k\boldsymbol{v}^T\boldsymbol{l}_c) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix} \tag{8}$$

where $q = 1 - k\boldsymbol{v}^T\boldsymbol{l}$, $s = -ka_v$ and $\mu = -kb_v$. The homography $\boldsymbol{H}_c$ has a very simple form and corresponding points, $\boldsymbol{x}_c$, $\boldsymbol{x}'_c$, in FOE centered images are related as follows:

$$\lambda \begin{bmatrix} x'_c \\ y'_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \tag{9}$$

If the robot's translation direction is parallel to the ground, $q = 1$ since the FOE lies on the vanishing line for this special motion. In this specialisation, the original H-matrix has four degrees-of-freedom, and is sometimes termed an elation (Hartley and Zisserman, 2001). Otherwise, the FOE is at a distance d_1 from the vanishing line and the five degree-of-freedom two-view planar relationship is termed a homology.

$$d_1 = \left| \frac{1 - q}{k\sqrt{a_v^2 + b_v^2}} \right| \tag{10}$$

Under the new homography, we have

$$(sx_c + \mu y_c + q)^2 \left(x_c'^2 + y_c'^2\right) = \left(x_c^2 + y_c^2\right) \tag{11}$$

If we define $\rho = \frac{1}{r}$, where $r = \sqrt{x_c^2 + y_c^2}$ is the Euclidean distance between an image point and the FOE in a frame, then taking square roots of Eq. (11)

$$f(\theta) = \rho' - q\rho = s\frac{x_c}{r} + \mu\frac{y_c}{r} = k_{s\mu}\sin(\theta + \alpha) \tag{12}$$

where θ is the angular position of a pixel in a frame centred on the FOE.

$$k_{s\mu} = \sqrt{s^2 + \mu^2}, \quad \sin\alpha = \frac{s}{k_{s\mu}}, \quad \cos\alpha = \frac{\mu}{k_{s\mu}}, \quad \tan\alpha = \frac{a_v}{b_v} \tag{13}$$

Eq. (12) indicates that we need to find three constants (q, $k_{s\mu}$, α) in order to recover the homography and that the computation should be implemented in (ρ, θ) image space (note that a planar homology has five dof, but two have been recovered in the FOE computation). We call this reciprocal-polar (RP) image space. Thus, after computing the FOE, an interpolation procedure is used to generate a RP image for each image in the image pair.

Along any radius from the FOE, $f(\theta)$ is constant, so for any two pairs of correspondences (i,j)

$$q = \frac{\rho'_i - \rho'_j}{\rho_i - \rho_j} \tag{14}$$

Values of ρ in above equation can be determined by 1D windowed correlations between the two images in RP image space, i.e., along lines of constant θ in the RP image space. q can be obtained by using all the strong correlation results.

Given a value q, the RP image of the first image may be scaled along the ρ dimension. This allows correlations to be made along the ρ dimension at each angle θ_i. For each pixel in the first image, its position in RP image space is computed, and a 1D window is created around this position along the $q\rho$ dimension. We then correlate this window along the ρ dimension in the second RP image 2, at the same value of θ. This correlation process is possible because of the 'pure-shift' relation between ρ' and $q\rho$, expressed in Eq. (12) and the position of the maximum value of the correlation is related as a value of $f(\theta)$.

Eq. (12) indicates that correlation maxima and feature correlations in reciprocal-polar space, which are associated with a planar surface, lie on a sinusoid and the constants ($k_{s\mu}$,α) may be recovered by fitting a sinusoid to the data for $f(\theta)$.

Suppose that we have two values of $f(\theta)$, $f_{i,j}$ measured at two angles, $\theta_{i,j}$, so that

$$f_i = k_{s\mu}\sin(\theta_i + \alpha), \quad f_j = k_{s\mu}\sin(\theta_j + \alpha) \tag{15}$$

$$\frac{f_i}{f_j} = \frac{\sin\theta_i + \cos\theta_i\tan\alpha}{\sin\theta_j + \cos\theta_j\tan\alpha} \tag{16}$$

collecting terms in $\tan\alpha$ and rearranging gives

$$\tan\alpha = \frac{f_j\sin\theta_i - f_i\sin\theta_j}{f_i\cos\theta_j - f_j\cos\theta_i} \tag{17}$$

Thus, in theory, a pair of f values, at different angular positions, for pixels belonging to the same plane, allows us to estimate the orientation of the vanishing line of that plane. Then, given the phase angle, α, corresponding to the orientation of the vanishing line, we can compute $k_{s\mu}$ from the Eq. (15).

In order to robustly and accurately estimate the vanishing line orientation from many correlation maxima and feature correspondences in reciprocal-polar space, many of which will not be associated with the ground plane, a RANdom SAmple Consensus (RANSAC) method (Fischer and Bolles, 1981) and iterated least-squares process are used. We define an optimization object function as:

$$\phi(\delta) = \sum_{i,j} \left| \delta - \frac{f_j\sin\theta_i - f_i\sin\theta_j}{f_i\cos\theta_j - f_j\cos\theta_i} \right| \tag{18}$$

where $\delta = \tan\alpha$, can be used to minimize $\phi(\delta)$, and determine the best set of inliers in the $f(\theta)$ data to a putative sinusoid. In this way, co-planar pixels may be grouped

without explicit construction of a homography matrix, although this is easily recovered in the FOE centred frame from the FOE and the two parameters of the sinusoid.

Let

$$f_i = k_{s\mu}\sin(\theta_i + \alpha) = s\cos\theta_i + \mu\sin\theta_i \tag{19}$$

Thus, for the all n inliers of the sinusoid model, we can write

$$\boldsymbol{AZ} = \boldsymbol{b} \tag{20}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_n & \sin\theta_n \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} s \\ \mu \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

$$\boldsymbol{Z} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{21}$$

The original homography matrix can be explicitly expressed as

$$\boldsymbol{H} = \boldsymbol{T}_{\mathrm{c}}^{-1}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix}\boldsymbol{T}_{\mathrm{c}} \tag{22}$$

From Eq. (6), we have

$$k\boldsymbol{v}\boldsymbol{l}^{\mathrm{T}} = \boldsymbol{I} - \boldsymbol{H} \tag{23}$$

If the jth row of the matrix $\boldsymbol{I} - \boldsymbol{H}$ is denoted by $\boldsymbol{h}^{j\mathrm{T}}$, then we may write

$$k\begin{pmatrix} x_f\boldsymbol{l}^{\mathrm{T}} \\ y_f\boldsymbol{l}^{\mathrm{T}} \\ \boldsymbol{l}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{h}^{1\mathrm{T}} \\ \boldsymbol{h}^{2\mathrm{T}} \\ \boldsymbol{h}^{3\mathrm{T}} \end{pmatrix} \tag{24}$$

Then we can get the vanishing line

$$\boldsymbol{l}^{\mathrm{T}} = \boldsymbol{h}^{3\mathrm{T}} \tag{25}$$

5. Affine height measurements

The approach described above allows pixels to be classified as either belonging to the reference plane or not. For those non-reference plane regions, we would like to know their height above the reference plane. Our aim is to compute the relative height, h_{r}, of corner point $\boldsymbol{A}$ in Fig. 2, as a fraction of the height, h_{c}, of the camera optical centre $\boldsymbol{O}$ above the reference (ground) plane, when the camera undergoes pure translation $\boldsymbol{t}$ and the motion direction is parallel to the reference plane. Point $\boldsymbol{A}$ is the actual position of the corner point relative to the camera before the translation and the $\boldsymbol{C}$ is the position of the corner after the translation. Points $\boldsymbol{A}'$ and $\boldsymbol{C}'$ are the projections of these actual corner positions onto the ground plane. Points $\boldsymbol{a}$ and $\boldsymbol{c}$ are the image positions of the corner at positions $\boldsymbol{A}$ and $\boldsymbol{C}$, respectively, and $\boldsymbol{b}$ is the predicted image position of the corner point by using H matrix of ground plane, if the corner point were to lie in the ground plane. Image point $\boldsymbol{b}$ is computed as $\boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}$, using the recovered homography.

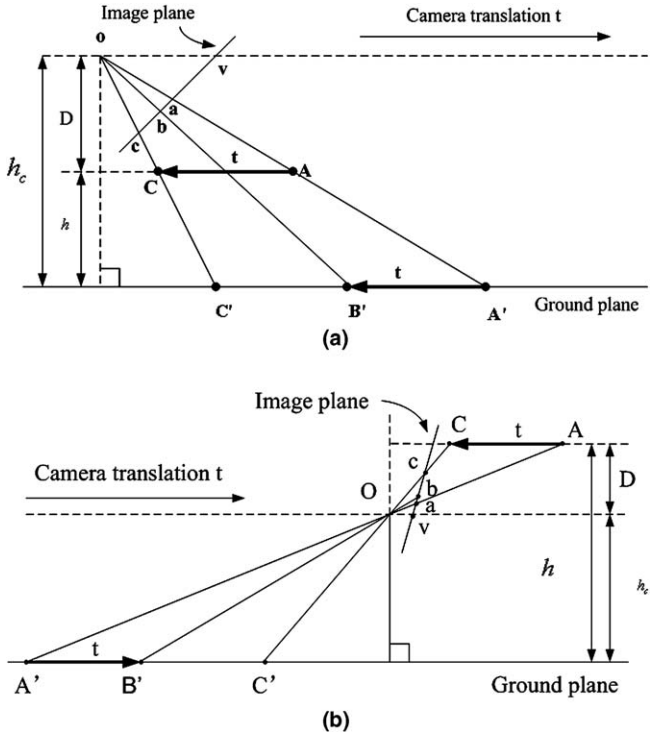


Fig. 2. Measuring the height of point $\boldsymbol{A}$. (a) $\boldsymbol{a}$ lies below the vanishing line and (b) $\boldsymbol{a}$ lies above the vanishing line.

Fig. 2(a) shows that point $\boldsymbol{a}$ lies below the vanishing line. Because $\boldsymbol{AC}//\boldsymbol{A}'\boldsymbol{C}'$, using similar triangles, and denoting the distance between point x and y as $d(x,y)$, we can get

$$h_{\mathrm{r}} = \frac{h}{h_{\mathrm{c}}} = 1 - \frac{D}{h_{\mathrm{c}}} = 1 - \frac{d(\boldsymbol{O},\boldsymbol{C})}{d(\boldsymbol{O},\boldsymbol{C}')} = 1 - \frac{d(\boldsymbol{A},\boldsymbol{C})}{d(\boldsymbol{A}',\boldsymbol{C}')} \tag{26}$$

For pure translation, $d(\boldsymbol{A},\boldsymbol{C}) = d(\boldsymbol{A}',\boldsymbol{B}')$, so that

$$h_{\mathrm{r}} = 1 - \frac{d(\boldsymbol{A}',\boldsymbol{B}')}{d(\boldsymbol{A}',\boldsymbol{C}')} \tag{27}$$

Now, the four image points $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{v})$, where $\boldsymbol{v}$ is the FOE, and their corresponding four ground plane points $(\boldsymbol{A}',\boldsymbol{B}',\boldsymbol{C}',\infty)$ are collinear. The cross-ratio for this set of points remains invariant under perspective projection transform and so we can get

$$\frac{d(\boldsymbol{A}',\boldsymbol{B}')}{d(\boldsymbol{A}',\boldsymbol{C}')} = \frac{d(\boldsymbol{a},\boldsymbol{b})d(\boldsymbol{c},\boldsymbol{v})}{d(\boldsymbol{a},\boldsymbol{c})d(\boldsymbol{b},\boldsymbol{v})} \tag{28}$$

Hence, the height of the corner point relative to the height of the optical centre is

$$h_{\mathrm{r}} = 1 - \frac{d(\boldsymbol{a},\boldsymbol{b})d(\boldsymbol{c},\boldsymbol{v})}{d(\boldsymbol{a},\boldsymbol{c})d(\boldsymbol{b},\boldsymbol{v})} \tag{29}$$

If the point $\boldsymbol{a}$ lies above the vanishing line (Fig. 2(b)), we have

$$h_{\mathrm{r}} = \frac{h}{h_{\mathrm{c}}} = 1 + \frac{D}{h_{\mathrm{c}}} = 1 + \frac{d(\boldsymbol{A},\boldsymbol{C})}{d(\boldsymbol{A}',\boldsymbol{C}')} \tag{30}$$

$$h_{\mathrm{r}} = 1 + \frac{d(\boldsymbol{a},\boldsymbol{b})d(\boldsymbol{c},\boldsymbol{v})}{d(\boldsymbol{a},\boldsymbol{c})d(\boldsymbol{b},\boldsymbol{v})} \tag{31}$$

If point $\boldsymbol{a}$ lies on the vanishing line, then $h_r = 1$.

The algorithm for computing the height of obstacles is summarised as follows:

$$h_r = \begin{cases} 1 - \dfrac{d(\boldsymbol{a},\boldsymbol{b})d(\boldsymbol{c},\boldsymbol{v})}{d(\boldsymbol{a},\boldsymbol{c})d(\boldsymbol{b},\boldsymbol{v})} & \boldsymbol{a} \text{ lies below the vanishing line} \\ 1 & \boldsymbol{a} \text{ lies on the vanishing line} \\ 1 + \dfrac{d(\boldsymbol{a},\boldsymbol{b})d(\boldsymbol{c},\boldsymbol{v})}{d(\boldsymbol{a},\boldsymbol{c})d(\boldsymbol{b},\boldsymbol{v})} & \boldsymbol{a} \text{ lies above the vanishing line} \end{cases} \tag{32}$$

Note that h_r can be interpreted as the height of point $\boldsymbol{A}$ in units of height h_c. The absolute distance can be obtained from this distance ratio once the camera's height h_c is specified. However, it is sometimes more practical to determine the distance via a second measurement in the image, that of a known reference length. Note that this approach only needs the H-matrix of ground plane, and the tracked image correspondences $\boldsymbol{a}$ and $\boldsymbol{c}$ of the feature to determine the height to the ground plane. The main advantages compared with other method (such as the method of Criminisi et al. (2000) and Wilczkowiak et al. (2001)) is that this method does not need any camera calibration or any geometry constraints in the scene, and it can be used to compute the height from any feature point to the reference plane.

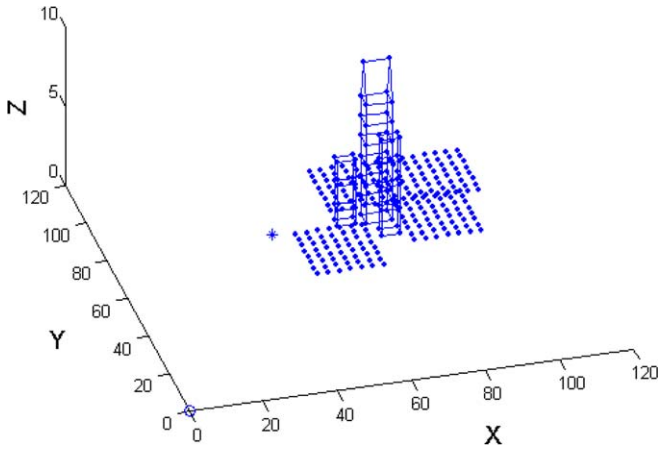


Fig. 3. A synthetic 3D scene.

6. Experimental results

A large number of synthetic data and real images were selected and intensive experimental work was carried out in order to test the robustness and accuracy of the method proposed in this paper.

6.1. Simulated experimental results

The simulated experiment was carried out on a 3D Euclidean model (Fig. 3). The intrinsic camera parameters were chosen as follows: the synthetic camera had a focal length which is invariant and an aspect ratio of one with no skew, $f/dx = f/dy = 500$, the principal point has a value of (319, 239) and the image size is 640×480. Fig. 4 shows the images taken from the camera's translation direction parallel to the ground and includes their FOEs (o) and vanishing lines.

One hundred and fifty matches on the floor are used to compute the $\boldsymbol{H}$ matrix. Note that we compared our method with the standard least-squares estimation of the planar

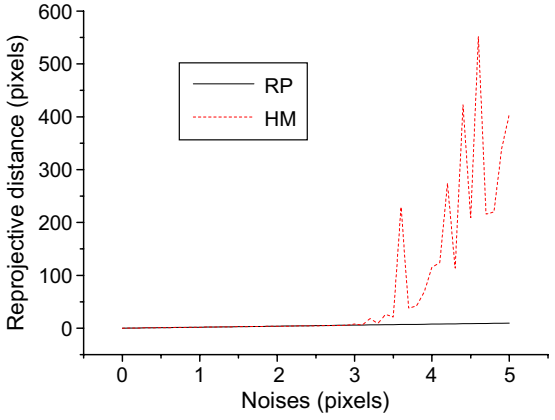


Fig. 5. Black line represents the reprojective errors of the RP method. Red dashed line represents the reprojective errors of the LS homography method (HM). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

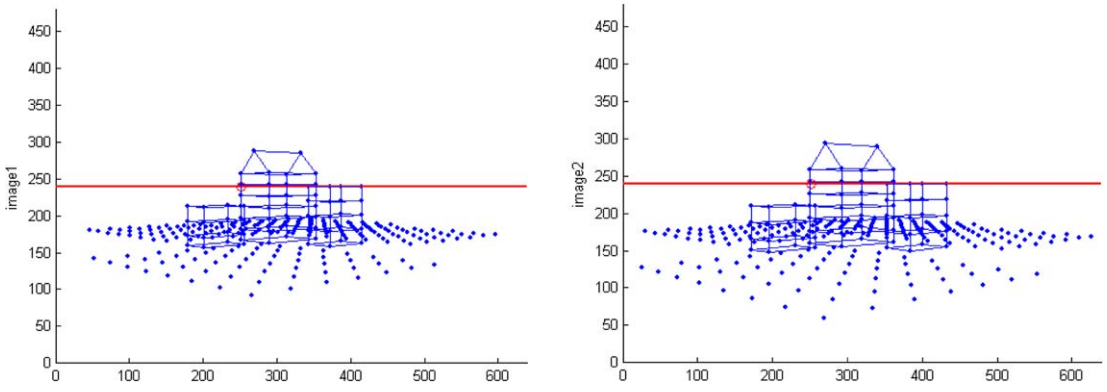
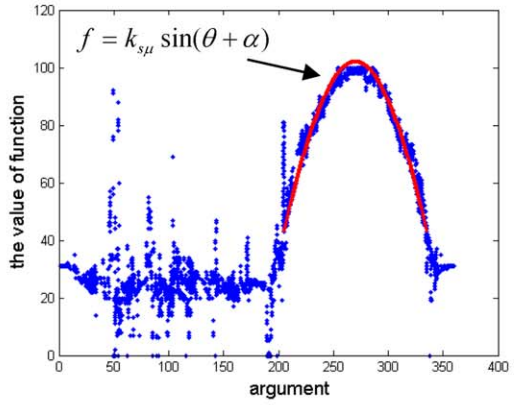


Fig. 4. Synthetic images used for simulation.

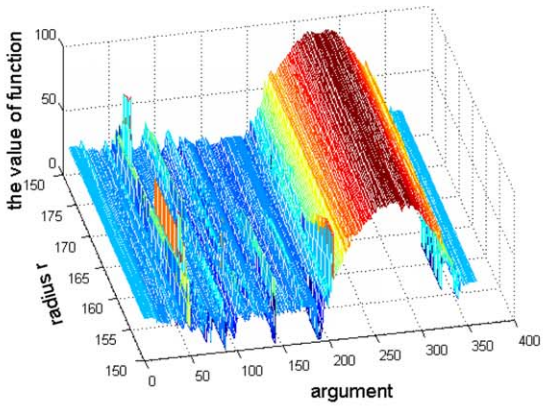
(a) (b)

(c)

(d)



(e)



(f)

Fig. 6. (a) One of the original images. (b) Another image with their matching points (●), displacement of feature points, FOE (o) and vanishing line. (c) and (d) The reciprocal-polar images corresponding to (a) and (b), respectively. (e) The value of $f(\theta)$. (f) 3D sinusoid.

homography matrix (Hartley and Zisserman, 2001). Different Gaussian noises with mean from 0 to 5 pixels, variance one and standard deviation one were added to the image points to test the robustness of the reciprocal-polar (RP) method. The program executed 1000 times for every noise level. The means of the results are shown in Fig. 5. The results show that, if noise level is less than 3 pixels, then the two methods have almost the same robustness and accuracy. However, when noise level is increased to more than 3 pixels, our RP method is more robust than the standard least-squares (LS) homography matrix estimation method. After the H-matrix was computed, the height of obstacles can be recovered by using Eq. (32).

6.2. Real images experimental results

In the two visual metrology experiments (one indoor, one outdoor) presented here, $q = 1$ was assumed since two VGA frames (640×480 resolution) were captured with the translation direction parallel to the ground plane, and in both cases, the height $\boldsymbol{AB}$, shown in Figs. 8 and 10, was used as the reference height.

The first experiment is conducted in an indoor scene, which resulted in 90.4% of intersections of all lines defined by all corner correspondences from the image pair to lie within a 50 pixels radius, which we classify as a (near) pure translation motion. The FOE (333.94, 264.41) was obtained using the robust estimation method presented here. Fig. 6(a) shows one of the sample image pair and Fig. 6(b) shows another image with their matching points (·), displacement of feature points, FOE (o) and vanishing line. After the FOE had been obtained, the images were then converted to RP (ρ, θ) form. The RP images are shown in Fig. 6(c) and (d). These are quite severe distortions from the original, since radial lines centred on the FOE in the original image are now horizontal scan lines, and the intensity variations along these scan lines are flipped and non-linearly distorted due to the effect of the reciprocal operation. The relationship between the two figures is that reference plane pixels have undergone a pure shift along horizontal scan lines and the magnitude of this shift follows a sinusoidal variation as we move down the image from one scan line to the next. Close inspection of the images does indeed reveal that this horizontal shifting process has occurred. Fig. 6(e) shows the reference plane image motion recovered using the 1D correlation method and the RANSAC method. The partial sinusoidal curve ($\theta \in [205°, 335°]$) is clearly shown which represents the motion of the ground plane in RP space. The phase of the curve is shown close to 180° rather than 0° because the direction of y in the image is directed upwards from the FOE rather than downwards. The three-dimensional sinusoidal form of RP ground plane motion, which is obtained when ρ is varied, is shown in Fig. 6(f). It clearly illustrates that the sinusoidal form is close to being constant irrespective of the radial distance of an image point from the FOE. In other words, planar image motion is a pure shift in RP space. Furthermore, we can get $\mu = -5.6374\text{e}-004$, and $\tan\alpha = 0$ and hence

$$\boldsymbol{H} = \begin{bmatrix} 1.0000 & -0.1883 & 49.7753 \\ 0 & 0.8509 & 39.4108 \\ 0 & -0.0006 & 1.1491 \end{bmatrix}$$

Fig. 7(a) and (b) show the reference plane (ground) recovered by the RP method and standard LS homography method, respectively. These results show the RP method performing better than the standard LS homography method.

Using the height of the box labelled "AB" (30 cm), we can recover the scale, after which we can correctly scale other heights in the scene, as shown in Fig. 8.

The second experiment implemented in an outdoor scene resulted in 89.9% of intersections of all lines defined by all corner correspondences from the image pair to lie within a 50 pixels radius after tracking the feature points. The FOE (261.17, 260.24) was obtained using the robust estimation method presented here. Fig. 9 shows one of the image pair with their matching points (·), displacement of feature points, FOE and vanishing line. We computed $s = -8.0429\text{e}-006$, $\mu = -2.9382\text{e}-004$ and $\tan\alpha = 0.0274$ and hence

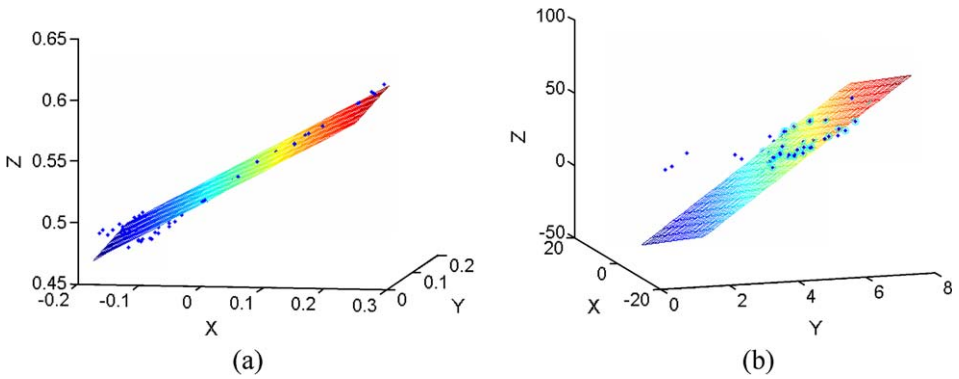


Fig. 7. (a) The reference plane (ground) recovered by RP and (b) the reference plane recovered by the standard LS homography method.

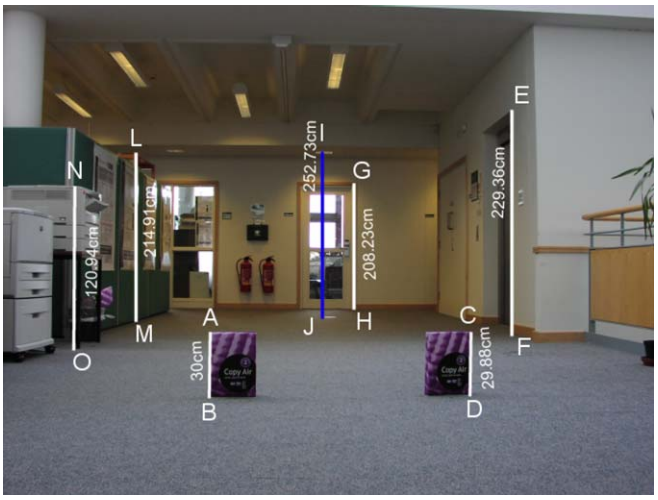


Fig. 8. Measuring the height of some interested point.

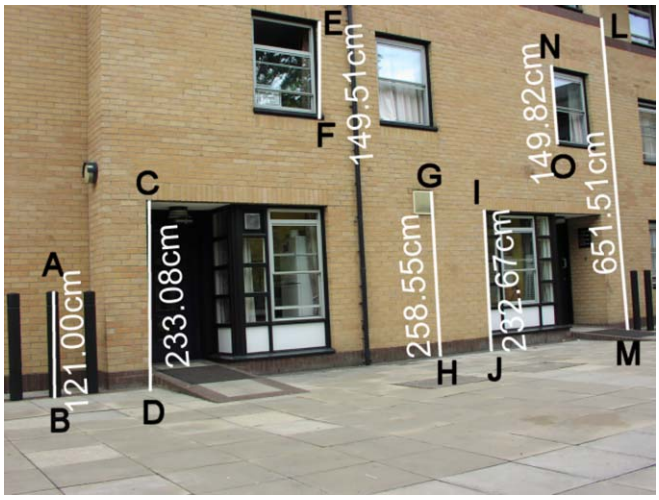


Fig. 10. Some measured results for outdoor scene.

Fig. 9. One of the image pair with their matching points (●), displacement of feature points, FOE (o) and vanishing line.

$$\boldsymbol{H} = \begin{bmatrix} 0.9979 & -0.0767 & 20.5190 \\ -0.0021 & 0.9235 & 20.4462 \\ 0 & -0.0003 & 1.0786 \end{bmatrix}$$

Fig. 10 shows the measured results. The quantitative results are given in Table 2, where 'TM' are the manual (tape measure) measurements, 'VM' are the visual metrology results, 'AE' are the absolute errors and 'RE' are the relative errors. We find a mean absolute error of 7.1 mm and mean relative error of 0.37%. If we remove the two rather inaccurate measurements EF and LM (indoor), the remaining measurements have a mean absolute error of 1.8 mm and a 0.13% mean relative error. We have also used the method presented in this paper to detect potential obstacles for robot navigation. In this application, we measure the height of contours around segmented image regions, and label them as obstacles if they are above a height threshold, set at a fraction of the robot wheel radius. Some of our obstacle detection results are shown in Fig. 11. Our system is coded in MATLAB and can run at around five frames per second on a PC with a Pentium IV processor running at 2.4 GHz and 512MB of memory. Finally note that our system performance is largely unaffected by the magnitude of the translation, with the proviso that sufficient image motion (more than a few pixels) is generated in order to detect that the camera motion is near pure translation.

Table 2
Visual metrology results in centimetres

Seg.	Indoor				Outdoor			
	TM	VM	AE	RE (%)	TM	VM	AE	RE (%)
CD	30.0	29.88	0.12	0.400	233.1	233.08	0.02	0.0086
EF	227.7	229.36	1.66	0.7290	149.8	149.51	0.29	0.1936
GH	208.4	208.23	0.17	0.0816	258.7	258.55	0.15	0.0580
IJ	252.5	252.73	0.23	0.0911	233.1	232.67	0.43	0.1845
NO	121.1	120.94	0.16	0.1321	149.8	149.82	0.02	0.0134
LM	210.3	214.91	4.61	2.1921	None	651.51		

Fig. 11. Obstacles detection results.

7. Conclusions

The main contributions in this work are (in the order of algorithm execution) (1) robust FOE estimation, (2) RP rectification, (3) planar image motion estimation, planar homography estimation and plane segmentation by robust estimation of a sinusoid in RP image motion space, and (4) a projective construction allowing affine height to a plane to be measured from an uncalibrated image pair. Intensive experimental work was carried out in order to test the ac-curacy of the method proposed in this paper. The results show that our algorithm performs very well to outliers and noise and provides a practical method of visual metrology.

Acknowledgement

The authors acknowledge the support of the UK DTI Aeronautics Research Programme.

References

Atkinson, K.B. (Ed.), 1996. Close Range Photogrammetry and Machine Vision. Whittles Publishing.

Criminisi, A., Reid, I., Zisserman, A., 1999. Aplane measuring device. Image Vision Comput. 17 (8), 625–634.

Criminisi, A., Reid, I., Zisserman, A., 2000. Single view metrology. Internat. J. Comput. Vision 40 (2), 123–148.

Faugeras, O., Luong, Q.-T., Papadopoulo, T., 2001. The Geometry of Multiple Images: The Laws that Govern the Formation of Multiple Images of a Scene and Some of their Applications. The MIT Press, Cambridge, Massachusetts, London, England.

Fischer, M.A., Bolles, R.C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. Assoc. Comput. Machine 24 (6), 381–395.

Forsyth, D.A., Ponce, Jean, 2003. Computer Vision: A Modern Approach. Prentice Hall, Upper Saddle River, NJ.

Harris, C.J., Stephens, M., 1988. A combined corner and edge detector. In 4th Alvey Vision Conference, Manchester, pp. 147–151.

Hartley, R., Zisserman, A., 2001. Multiple View Geometry in Computer Vision. Cambridge University Press (reprinted).

Heyden, A., Berthilsson, R., 1999. An iterative factorization method for projective structure and motion from image sequences. Image Vision Comput. 17 (13), 981–991.

Kim, T., Seo, Y., Hong, K., 1998. Physics-based 3D position analysis of a soccer ball from monocular image sequences. Proc. ICCV, 721–726.

Liang, B., Pears, N.E., 2002a. Ground plane segmentation from multiple visual cues. Second Internat. Conf. on Image and Graphics, Hefei, China, Proc. of SPIE, vol. 4875, pp. 822–829.

Liang, B., Pears, N.E., 2002b. Visual navigation using planar homographies. In: Proc. of IEEE Internat. Conf. on Robotics and Automation. Washington, DC, USA, vol. 1, pp. 205–210.

Pollefeys, M., Koch, R., Vergauwen, M., Gool, L.Van., 1998. Metric 3D surface reconstruction from uncalibrated image sequences. In: Proc. SMILE Workshop (Post-ECCV'98), LNCS1506. Springer-Verlag, pp. 138–153.

Reid, I., Zisserman, A., 1996. Goal-directed Video Metrology. In: Cipolla, R., Buxton, B. (Eds.), Proc. ECCV, vol. II. Springer, pp. 647–658.

Shi, Jianbo, Tomasi, Carlo, 1994. Good features to track. IEEE Conf. on Computer and Pattern Recognition, Seattle, USA, pp. 593–600.

Smith, S., Brady, J., 1997. SUSAN-a new approach to low level image processing. Internat. J. Comput. Vision 23 (1), 45–78.

Tomasi, C., Kanade, T., 1992. Shape and motion from image streams under orthography: A factorization approach. Internat. J. Comput. Vision 9 (2), 137–154.

Triggs, B., 1996. Factorization methods for projective structure and motion. In: Proc. Conf. on Computer Vision and Pattern Recognition, pp. 845–851.

Wilczkowiak, M., Boyer, E., Sturm, P.F., 2001. Camera calibration and 3d reconstruction from single images using parallelepipeds. Internat. Conf. on Computer Vision, Vancouver, pp. 142–148.

Zhang, Z., 1998. Determining the epipolar geometry and its uncertainty: a review. Internat. J. Comput. Vision 27 (2), 161–195.