Projective rectification based on relative modification and size extension for stereo image pairs

H.-H.P. Wu, Y.-H. Yu and W.-C. Chen

Abstract: For stereo vision applications, projective geometry has proved to be a useful tool for solving the rectification problem without camera calibration. However, the criterion of minimisation for projective rectification must be chosen properly in order to avoid unduly geometric distortion. In this paper, an improved algorithm to minimise the distortion by combining a newly developed projective transform with a properly chosen shearing transform is proposed. The emphasis on low geometric distortion makes this method not only appropriate for 3-D reconstruction but also for stereoscopic viewing applications. On the basis of relative modification, this new method contains fewer parameters (6 degrees of freedom) for minimisation, which reduces the processing time, and improves the rectification result. Several different types of image pairs were tested to demonstrate the applicability and reliability of the proposed algorithm visually and quantitatively. Comparisons with other methods are also provided to verify the improvement of this new scheme.

1 Introduction

Stereopsis is the most important cue for sense of depth for both humans and machines. It is created by analysing two views jointly to produce 3-D information. The most fundamental step in this stereo reconstruction process is to establish correspondence between the two images. The correspondence problem is defined as locating a pair of image pixels from two different images, where these two pixels are projections of the same scene element. For the human visual system, this can be accomplished naturally; however, for 3-D machine vision, this is a time-consuming computational process. Given a point in one image, its correspondent point (or point correspondence) must lie on an epipolar line in the other image. This relationship is well known as the epipolar constraint [1]. It is obvious that knowledge of this epipolar geometry, or codified as fundamental matrix, simplifies the stereo matching from a 2-D area search to a 1-D search along the epipolar line. If the images are acquired from a pair of identical cameras placed side-by-side and pointed in the same direction, known as a rectilinear stereo rig, the epipolar lines will coincide with scan lines (**x**-axis) of the images. Given this ideal epipolar geometry, the correspondent points will lie on the same scan line in the two images. However, for an arbitrary placement of cameras, the epipolar lines are skew and the 1-D search will still be time consuming.

When the epipolar geometry is not in an ideal form, the image pairs can be warped to make correspondent points lie on the same scan lines. This process is known as image rectification, and can be accomplished by applying 2-D projective transforms, or homographies, to each image. The homography is a linear one-to-one transformation of the projective plane, which is represented by a 3×3 non-singular matrix. The rectified images can then be treated as obtained by a rectilinear stereo rig and the correspondence problem is greatly simplified. Since most stereo algorithms assume input images having ideal epipolar geometry, image rectification is usually a pre-requisite operation for stereoscopic related applications.

The idea of rectification has long been used in photogrammetry [2]. The techniques originally used were optical based, but now are replaced by software methods that model the geometry of optical projection. The software-based photogrammetric approaches, similar to most of the computer vision ones, assume the knowledge of projection matrices or cameras parameters [3–5]. These methods require camera parameters to compute a pair of homographies for transformations. The necessity of camera calibration is one of their disadvantages.

In contrast to these traditional approaches, several researchers have developed techniques called projective rectification to rectify images directly without using camera parameters [6–13]. They utilised the epipolar geometry of the acquired images and various criteria to compute the homographies. Robert *et al.* [6] attempted to find the transform that best preserves orthogonality around image centres. Hartley [7] proposed using minimisation of the differences between matching points for the solution of homographies. He also gave a detailed theoretical presentation of the projective rectification. Loop and Zhang [8] suggested decomposing each homography into projective and affine components. They then found the projective component that minimises a defined projective distortion criterion. Pollefeys [9] proposed a simple and efficient algorithm for general two view stereo images. Although this method can obtain a minimal image size without any pixel loss, it requires projection matrix data. Chen [10] modified Pollefeys's method and replaced the projection matrix with an F matrix to compute the homographies. Given the fundamental matrix and epipoles, Al-Shalfan [11]

IEE Proceedings online no. 20045002

doi: 10.1049/ip-vis:20045002

Paper first received 19th April 2004 and in revised form 21st February 2005. Originally published online 8th June 2005

H.-H.P. Wu and W.-C. Chen are with the Department of Electrical Engineering, National Yunlin University of Science and Technology, #123 University Road, Section 3, Douliu 640, Taiwan, ROC,

Y.-H. Yu is with the Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, #123 University Road, Section 3, Douliu 640, Taiwan, ROC, and is also with the Department of Computer Science and Information Engineering, Nan Kai Institute of Technology, Taiwan, ROC

E-mail: wuhp@yuntech.edu.tw

suggested a direct approach to the rectification problem based on singular value decomposition. While these proposed methods provided many possibilities for projective rectification, they must explicitly estimate the fundamental matrix before rectification and can only solve the problem indirectly. Since there are many ways of obtaining the fundamental matrix, which generate many possible solutions [15], this indirect approach might obtain varying rectified results.

Isgrò and Trucco [13] were the first ones to obtain homographies directly without computing the fundamental matrix. Let us refer to this method as the 'IT method' in the following discussion. While the IT method simplifies the problem of rectification, it requires disparity minimisation along the **x**-axis to generate a unique solution. In certain applications (e.g. 3-D reconstruction), this modification of the **x**-axis disparity in rectification might be harmless; however, in applications where original **x**-axis disparity must be maintained (e.g. for stereoscopic viewing purposes), this constraint will make the algorithm useless. Moreover, the enforcement of minimising **x**-axis disparity to obtain a single solution sometimes greatly distorts the image. Owing to the popularity and convenience of digital cameras, one can now effortlessly obtain 3-D image pairs and view them on an autostereoscopic LCD display (e.g. DTI [17]) or by shutter glasses (e.g. CrystalEyes liquid crystal shutter eyewear [18]). Since 3-D can provide more realistic perception, it has been used in entertainment, gaming, simulation, teleconferencing, and tele-operation to augment the reality of presence. One prominent application of this 3-D technology is the European IST project VIRTUE [19, 20], which aims to produce an impression of presence in an immersive teleconferencing system. Therefore, the availability of cheap digital stereoscopic viewing devices will make applications of stereo images more popular. The stereo image pair can be acquired by simply shooting the common scene at different positions using a single camera. Obviously, image pairs taken this way will not satisfy the condition of the rectilinear stereo rig and cannot be viewed directly to perceive 3-D. Fortunately, the image rectification technique proposed in this paper can be used to warp the image pair and make the perception of 3-D easy and possible.

In this paper, we propose a new approach for rectifying two uncalibrated images with reduced geometric distortion. The emphasis on low geometric distortion allows our algorithm to be used not only for stereo vision, but also for stereoscopic viewing purposes. Its novelty lies in formulating a new set of parameters using the concept of relative modification. This leads to reduced degrees of freedom for homographies in solving the rectification problem. The proposed approach possesses similar advantages to that of the IT method by performing uncalibrated rectification without explicit computation of the epipolar geometry (fundamental matrix). Furthermore, the new method combines a novel image size pre-extension scheme and a shearing transform, which greatly reduces geometric distortion caused by resampling in rectification.

2 Epipolar geometry

The derivation of our algorithm is presented from the viewpoint of projective geometry [1]. The image point is expressed in homogeneous coordinates and represented by a 3-dimensional column vector. In this paper column vectors are denoted by bold lower-case letters, such as $\mathbf{m}$ and $\boldsymbol{l}$. Matrices are represented by bold upper-case letter, such as $\mathbf{F}$ and $\mathbf{H}$. Transposed vectors and matrices are expressed by adding a letter T as a superscript, e.g. $\mathbf{m}^T$ and $\mathbf{F}^T$.

2.1 Epipolar constraint

Consider two images I and I' of a common scene. Let C and C' represent the optical centres of the left and right cameras in the 3-D coordinates, respectively. Points $\mathbf{m}$ and $\mathbf{m}'$ are the projections of a certain 3-D point M on the left (I) and right (I') images, as shown in Fig. 1. Two epipoles of the left and right images are represented by $\mathbf{e}$ and $\mathbf{e}'$, respectively. Given a fundamental matrix $\mathbf{F}$ and two corresponding points $\mathbf{m}$ and $\mathbf{m}'$, the epipolar constraint [1] asserts that

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

The matrix $\mathbf{F}$ is a 3×3 matrix with rank 2, and the epipoles for the left ($\mathbf{e} \in I$) and the right image ($\mathbf{e}' \in I'$) satisfy

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

Furthermore, all the epipolar lines ($\mathbf{l}$ and $\mathbf{l}'$) will pass the epipoles, or

$$\mathbf{l}^T\mathbf{e} = \mathbf{l}'^T\mathbf{e}' = 0 \tag{3}$$

2.2 Epipolar geometry after stereo image rectification

Image rectification can be treated as a process of converting the epipolar geometry of an image pair into a canonical form. This canonical epipolar geometry satisfies the following two properties: (1) all epipolar lines are parallel to the **x**-axis, and (2) corresponding points have identical y-coordinates. The rectification will map the epipole to a point at infinity. We designate these two epipoles after rectification as $\mathbf{e}_\infty$ and $\mathbf{e}'_\infty$, where $\mathbf{e}_\infty = \mathbf{e}'_\infty = [1\ 0\ 0]$, and the fundamental matrix of a pair of rectified images has the form of

$$\mathbf{F}_\infty = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{4}$$

Let $(\tilde{\mathbf{m}},\tilde{\mathbf{m}}')$ be an image pair in the rectified images corresponding to the original $(\mathbf{m},\mathbf{m}')$ pair. From (1), the epipolar constraint after rectification can be rewritten as

$$\tilde{\mathbf{m}}'^T\mathbf{F}_\infty\tilde{\mathbf{m}} = 0 \tag{5}$$

Meanwhile, assume that the rectification can be accomplished by the following operations:

$$\tilde{\mathbf{m}} = \mathbf{H}\mathbf{m},\ \ \tilde{\mathbf{m}}' = \mathbf{H}'\mathbf{m}' \tag{6}$$

where $\mathbf{H}$ and $\mathbf{H}'$, are the rectifying homographies (or H matrices) for the left and right images, respectively. Combining (5) and (6) we obtain the following equation:

$$\mathbf{m}'^T\mathbf{H}'^T\mathbf{F}_\infty\mathbf{H}\mathbf{m} = 0 \tag{7}$$

Given several point correspondences, or $(\mathbf{m}, \mathbf{m}')$ pairs, (7) can be solved to obtain homographies ($\mathbf{H}$ and $\mathbf{H}'$). This and other rectification issues are the topics of the next Section.

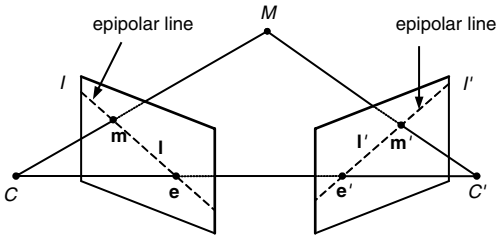

Fig. 1 *Epipolar geometry of a pair of stereo images*

3 Image rectification

Several projective rectification methods have been proposed recently, and the background of the methods that are most closely related to our algorithm will be briefly described below. On the basis of these works, we propose a new approach to the projective rectification problem.

3.1 Proposed method and F matrix parameterisation

In order to constrain the geometric distortion caused by rectification, Hartley [7] proposed that one of the two homographies, say $\mathbf{H}'$, should be close to a rigid transformation in the neighborhood of a selected point $\mathbf{p}_0$. That is, the homography for one of the images (I') can be represented by

$$\mathbf{H}' = \mathbf{KRT} \tag{8}$$

where $\mathbf{T}$ is a translational vector taking $\mathbf{p}_0$ to the origin, $\mathbf{R}$ is a rotation matrix mapping the epipole to a point $[1 \quad 0 \quad f]^T$ on the $\mathbf{x}$-axis, and $\mathbf{K}$ is a transformation matrix mapping $[1 \quad 0 \quad f]^T$ to a point $[1 \quad 0 \quad 0]^T$ at infinity along the $\mathbf{x}$-axis. Moreover, matrix $\mathbf{K}$ can be expressed as

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -f & 0 & 1 \end{bmatrix} \tag{9}$$

In this way, $\mathbf{H}'$ depends only on two parameters: f and rotation angle θ. If the translational vector $\mathbf{T}$ is neglected, then $\mathbf{H}'$ becomes

$$\mathbf{H}' = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ -f\cos\theta & -f\sin\theta & 1 \end{bmatrix} \tag{10}$$

Differing from Hartley's approach, we adopt a more straightforward form for $\mathbf{H}$ and $\mathbf{H}'$, as will be described below. Surprisingly, this more direct parameterisation scheme for homographies will lead to better rectifying results.

Given an originally-ideal camera arrangement, what kind of changes to the camera position would break the condition of canonical epipolar geometry? Obviously, the available modifications are: translation, zooming, rotation, panning, and tilting. The rectification can be treated as reversing the above process by transforming one pair of images to a common coordinate system such that the epipolar lines will be made parallel to the image $\mathbf{x}$-axis. The reverse effect can be achieved by the following projective transformations [16] (up to a scale factor):

$$\textit{Translation} : P_T^l = \begin{bmatrix} 1 & 0 & a_3 \\ 0 & 1 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

$$\text{Zooming} : P_T^z = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

$$\text{Rotation} : P_T^r = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

$$\text{Panning} : P_T^p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_7 & 0 & 1 \end{bmatrix} \tag{14}$$

$$\text{Tilting} : P_T^t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & a_8 & 1 \end{bmatrix} \tag{15}$$

As shown in Fig. 2, the translation and zooming account for the camera shifting in the $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ directions. The rotation, panning and tilting reverse the camera rotation around the $\mathbf{z}$, $\mathbf{x}$ and $\mathbf{y}$ axes, respectively. Applying a proper combination of the above transformations to the image pair can achieve the goal of rectification. Obviously, modification of the image pair is a relative one; for example, we can either scale one up or scale the other one down (Fig. 2*b*), and there is no need to apply the same type of transform to both images. The only exception is the panning P_T^p, since we need it for both images in order to transform the epipoles to infinity. The rectification is divided into two stages; the $\mathbf{y}$-axis disparity is minimised first, and then a shearing transform will be applied to minimise the geometric distortion.

On the basis of the above analysis and the concept of relative modification, we can attribute the rotation and panning to the right image and obtain

$$\begin{aligned} \mathbf{H}' &= P_T^p P_T^r \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_7 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ a_7\cos\theta & a_7\sin\theta & 1 \end{bmatrix} \end{aligned}$$

Replacing a_7 with $-f$ in the above equation yields

$$\mathbf{H}' = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ -f\cos\theta & -f\sin\theta & 1 \end{bmatrix} \tag{16}$$

The resulting $\mathbf{H}'$ equals the one defined in (10) and confirms Hartley's argument. Let the left image account for the translation, zooming, tilting and panning. Then we have

$$\begin{aligned} \mathbf{H} &= P_T^p P_T^t P_T^z P_T^l \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_7 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & a_8 & 1 \end{bmatrix} \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & a_3 \\ 0 & 1 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} a_1 & 0 & a_1a_3 \\ 0 & a_5 & a_5a_6 \\ a_1a_7 & a_5a_8 & \kappa \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \end{aligned} \tag{17}$$

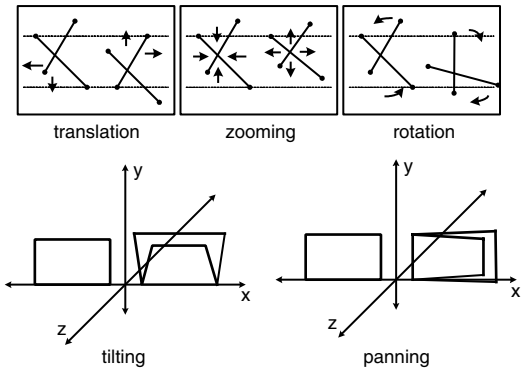

Fig. 2 *Relative modification of projective transforms*

where $\kappa = a_1a_3a_7 + a_5a_6a_8 + 1$, and all items are normalised by κ to obtain h_i. Because h_3 only affects x-coordinates and has no effect on y-disparity, we will leave it out in the process of rectification and obtain

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 & 0 \\ 0 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \tag{18}$$

Since the homography pair, $\mathbf{H}$ and $\mathbf{H}'$, are determined up to a scale factor, we have $\mathbf{H}(3,3) = \mathbf{H}'(3,3) = 1$. Combining (5), (6), and (7) gives us

$$\begin{aligned} \bar{\mathbf{m}}'^T\mathbf{F}_\infty\bar{\mathbf{m}} &= (\mathbf{H}'\mathbf{m}')^T\mathbf{F}_\infty(\mathbf{H}\mathbf{m}) = \mathbf{m}'^T(\mathbf{H}'^T\mathbf{F}_\infty\mathbf{H})\mathbf{m} \\ &= \mathbf{m}'^T\mathbf{F}\mathbf{m} = \mathbf{0} \end{aligned} \tag{19}$$

where $\mathbf{F} = \mathbf{H}'^T\mathbf{F}_\infty\mathbf{H}$. Substituting$\mathbf{F}_\infty$, $\mathbf{H}'$, and $\mathbf{H}$ in (4), (16) and (18) into $\mathbf{F}$, we can parameterise the F matrix as follows:

$$\begin{aligned} \mathbf{F} &= \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ -f\cos\theta & -f\sin\theta & 1 \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} h_1 & 0 & 0 \\ 0 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \\ &= \begin{bmatrix} h_7\sin\theta & -h_5f\cos\theta + h_8\sin\theta & -h_6f\cos\theta + \sin\theta \\ -h_7\cos\theta & -h_5f\sin\theta - h_8\cos\theta & -h_6f\sin\theta - \cos\theta \\ 0 & h_5 & h_6 \end{bmatrix} \end{aligned} \tag{20}$$

Obviously, both the F matrix and H matrix can be determined by the six parameters shown in vector form

$$\mathbf{\Phi} = [f \quad \theta \quad h_5 \quad h_6 \quad h_7 \quad h_8]^T \tag{21}$$

Note that we have combined the problems of rectification and estimation of the F matrix. In the next Section, we will show how to derive a least squares solution for the rectification problem. This parameterisation scheme combining with a shearing transform introduced later, leads to a unique solution for rectification.

3.2 Projection rectification based on least squares distance

In optimisation for rectification, we would like the criterion to be something geometrically meaningful and to be measurable in the image plane. One such quantity is the distance from a point $\mathbf{m}'_i$ to its corresponding epipolar line $\mathbf{l}'_i = \mathbf{F}\mathbf{m}_i = [l'_1 \quad l'_2 \quad l'_3]^T$, as shown in Fig. 3. This distance is given by the following equation:

$$d(\mathbf{m}'_i, \mathbf{l}'_i) = d(\mathbf{m}'_i, \mathbf{F}\mathbf{m}_i) = \frac{\mathbf{m}'^T_i\mathbf{l}'_i}{\sqrt{l'^2_1 + l'^2_2}} = \frac{\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i}{\sqrt{l'^2_1 + l'^2_2}} \tag{22}$$

Accordingly, distance for a point $\mathbf{m}_i$ to its corresponding epipolar line $\mathbf{l}_i = \mathbf{F}^T\mathbf{m}'_i = [l_1 \quad l_2 \quad l_3]^T$ is

$$d(\mathbf{m}_i, \mathbf{l}_i) = d(\mathbf{m}_i, \mathbf{F}^T\mathbf{m}'_i) = \frac{\mathbf{m}^T_i\mathbf{l}_i}{\sqrt{l^2_1 + l^2_2}} = \frac{\mathbf{m}^T_i\mathbf{F}^T\mathbf{m}'_i}{\sqrt{l^2_1 + l^2_2}} \tag{23}$$

We now demonstrate how to find the solution by minimising the distance defined above. To prevent inconsistency of the epipolar geometry between the left and right images, we choose to operate simultaneously on both images and minimise the mean-square distance. Hence, the problem becomes

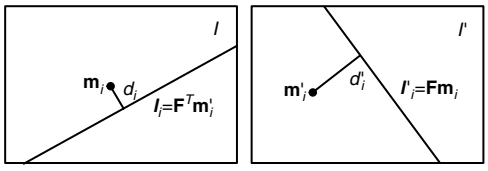

Fig. 3 *Distance from a point to the epipolar line of its correspondent point*

$$\min_{\mathbf{F}} \sum_i \frac{1}{2}\left(d^2(\mathbf{m}'_i, \mathbf{F}\mathbf{m}_i) + d^2(\mathbf{m}_i, \mathbf{F}^T\mathbf{m}')\right) \tag{24}$$

Using (22), (23) and the fact that $\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i = \mathbf{m}^T_i\mathbf{F}^T\mathbf{m}'_i$, we reformulate the minimisation problem in (24) as:

$$\begin{aligned} \min_{\mathbf{F}} \sum_i \frac{1}{2}\left[\frac{(\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i)^2}{l'^2_1 + l'^2_2} + \frac{(\mathbf{m}^T_i\mathbf{F}^T\mathbf{m}'_i)^2}{l^2_1 + l^2_2}\right] \\ = \min_{\mathbf{F}} \sum_i w_i \frac{(\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i)^2}{2} \end{aligned} \tag{25}$$

where $w_i = \dfrac{1}{l^2_1 + l^2_2} + \dfrac{1}{l'^2_1 + l'^2_2}$.

Given N point correspondences from the image pair ($\mathbf{m}_i \leftrightarrow \mathbf{m}'_i$, $i = 1 \ldots N$), the search for parameter vector $\mathbf{\Phi}$, which is to be used in $\mathbf{F} = \mathbf{F}(\mathbf{\Phi})$, $\mathbf{H} = \mathbf{H}(\mathbf{\Phi})$, and $\mathbf{H}' = \mathbf{H}'(\mathbf{\Phi})$, becomes a nonlinear optimisation problem, that is

$$\min_{\mathbf{F}} \frac{1}{N}\sum_{i=1}^{N} w_i \frac{(\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i)^2}{2} \tag{26}$$

To simplify the derivation, we can restructure the matrix equation by turning a matrix into a vector. If the 3×3 fundamental matrix in vector form is $\mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2 \quad \mathbf{F}_3]$, then we can use a **vec** operator to convert the matrix $\mathbf{F}$ into a column vector $\mathbf{f}$ by stacking the columns of $\mathbf{F}$, or

$$\mathbf{f} = \mathbf{vec}(\mathbf{F}) = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \mathbf{F}_3 \end{bmatrix} \tag{27}$$

Let symbol $\otimes$ denote the Kronecker product; then

$$\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i = (\mathbf{m}'_i \otimes \mathbf{m}_i)^T\mathbf{f} \tag{28}$$

Substituting (28) into (26) yields the objective function in vector form. That is

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} w_i \frac{(\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i)^2}{2} &= \frac{1}{2N}\sum_{i=1}^{N}\left(w_i^{1/2}\mathbf{m}'^T_i\mathbf{F}\mathbf{m}_i\right)^2 \\ &= \frac{1}{2N}\left\|\mathbf{W}^{1/2}\mathbf{U}_N\mathbf{f}\right\|^2 \end{aligned} \tag{29}$$

where $\mathbf{W}$ is an $N \times N$ diagonal matrix and $\mathbf{U}_N$ is an $N \times 9$ matrix:

$$\begin{aligned} \mathbf{W} &= \mathrm{diag}[w_1 \quad w_2 \quad \cdots \quad w_N] \\ &= \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & w_N \end{bmatrix}, \quad \mathbf{U}_N = \begin{bmatrix} (\mathbf{m}'_1 \otimes \mathbf{m}_1)^T \\ (\mathbf{m}'_2 \otimes \mathbf{m}_2)^T \\ \vdots \\ (\mathbf{m}'_N \otimes \mathbf{m}_N)^T \end{bmatrix}. \end{aligned}$$

Given the new notation, minimisation of the mean-square distance in (26) can now be rewritten as

$$\min_{\mathbf{f}} \frac{1}{2N} \left\| \mathbf{W}^{1/2}\mathbf{U}_N\mathbf{f} \right\|^2 \tag{30}$$

The solution for $\mathbf{f}$ in (30) can be found by any standard nonlinear minimisation method. We chose the Levenberg-Marquardt algorithm because of it effectiveness and popularity. Before applying this minimisation process, we need to derive the Jacobian matrix of (29), or $\frac{\partial}{\partial \mathbf{\Phi}}(\mathbf{W}^{1/2}\mathbf{U}_N\mathbf{f})$. Since factor $\mathbf{W}^{1/2}\mathbf{U}_N$ inside the partial derivative can be treated as a constant to $\mathbf{\Phi}$, the Jacobian matrix can be reduced to a simpler form

$$\frac{\partial}{\partial \mathbf{\Phi}}(\mathbf{W}^{1/2}\mathbf{U}_N\mathbf{f}) = (\mathbf{W}^{1/2}\mathbf{U}_N)\frac{\partial}{\partial \mathbf{\Phi}}\mathbf{f} \tag{31}$$

Therefore, we only need to compute $\frac{\partial}{\partial \mathbf{\Phi}}\mathbf{f}$ for the Jacobian matrix of (29) as follows

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{\Phi}}\mathbf{f} &= \frac{\partial}{\partial \mathbf{\Phi}} \begin{bmatrix} h_7\sin\theta \\ -h_7\cos\theta \\ 0 \\ -h_5 f\cos\theta + h_8\sin\theta \\ -h_5 f\sin\theta - h_8\cos\theta \\ h_5 \\ -h_6 f\cos\theta + \sin\theta \\ -h_6 f\sin\theta - \cos\theta \\ h_6 \end{bmatrix} \\
&= \begin{bmatrix} 0 & h_7\cos\theta & 0 \\ 0 & h_7\sin\theta & 0 \\ 0 & 0 & 0 \\ -h_5\cos\theta & h_5 f\sin\theta + h_8\cos\theta & -f\cos\theta \\ -h_5\sin\theta & -h_5 f\cos\theta + h_8\sin\theta & -f\sin\theta \\ 0 & 0 & 1 \\ -h_6\cos\theta & h_6 f\sin\theta + \cos\theta & 0 \\ -h_6\sin\theta & -h_6 f\cos\theta + \sin\theta & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&\quad \times \begin{bmatrix} 0 & \sin\theta & 0 \\ 0 & -\cos\theta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sin\theta \\ 0 & 0 & -\cos\theta \\ 0 & 0 & 0 \\ -f\cos\theta & 0 & 0 \\ -f\sin\theta & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}
\end{aligned} \tag{32}$$

The Jacobian matrix can be obtained by substituting (32) into (31), which is then used in an iterative process using the Levenberg-Marquardt algorithm to find the parameter vector $\mathbf{\Phi}$. This minimisation algorithm contains an iterative process, which must start with an initial estimate of the F matrix, and the ideal F matrix of a pair of rectified images can be used, that is,

$$\mathbf{F}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \mathbf{F}_\infty$$

A comparison of $\mathbf{F}_0$ with the parameterisation of $\mathbf{F}$ in (17) shows that $\mathbf{F}_0$ corresponds to an initial parameter vector of $\mathbf{\Phi}_0 = [\,1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0\,]^T$. An iterative process can now be applied to find the solution of $\mathbf{\Phi}$ after proper stop conditions have been set.

3.3 Geometric distortion reduction by shearing transform

After the parameter vector $\mathbf{\Phi} = [\,f \quad \theta \quad h_5 \quad h_6 \quad h_7 \quad h_8\,]^T$ has been found, the values of the vector can be used in (16) and (18) to calculate the pair of rectifying homographies shown as below

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 & 0 \\ 0 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix}, \quad \mathbf{H}' = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ -f\cos\theta & -f\sin\theta & 1 \end{bmatrix} \tag{33}$$

Obviously, parameter h_1 has not been estimated in the above minimisation process; therefore, the solution is actually not unique. Since h_1 does not affect the coordinate of the **y**-axis, we can simply set it to any value and still obtain rectifying results with approximately parallel epipolar lines. However, extra constraint on the transformed coordinate of the **x**-axis can be applied to obtain different values of $[\,h_1 \quad h_2 \quad h_3\,]$ for improving image quality. One suitable constraint is to keep the aspect ratio and perpendicularity of the central axes in the original images invariant after rectification. We will adopt the idea of the shearing transform described in [8] to achieve this purpose, and the procedure will be stated below. As will be shown subsequently, this approach can effectively reduce the overall geometric distortion.

Assume that the parameter vector $\mathbf{\Phi} = [\,f \quad \theta \quad h_5 \quad h_6 \quad h_7 \quad h_8\,]^T$ has been found by following the procedure described in the previous Section. Let $[\,h_1 \quad h_2 \quad h_3\,] = [\,1 \quad 0 \quad 0\,]$; then we have a preliminary solution of homographies

$$\mathbf{H}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix}, \quad \mathbf{H}'_0 = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ -f\cos\theta & -f\sin\theta & 1 \end{bmatrix}$$

To preserve the aspect ratio and perpendicularity of the axes after rectification, these two homographies can further be combined with the shearing transform defined below

$$\mathbf{H}_s = \begin{bmatrix} a & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{H}'_s = \begin{bmatrix} a' & b' & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{34}$$

Although it is not possible to remove the deformation caused by the panning parameter in the projective transform completely, this shearing transform can reduce the geometric distortion in the x-coordinate. The final homographies for the rectification will be a combination of two transforms:

$$\mathbf{H} = \mathbf{H}_s\mathbf{H}_0 = \begin{bmatrix} a & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{H}_0, \mathbf{H}' = \mathbf{H}'_s\mathbf{H}'_0 = \begin{bmatrix} a' & b' & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{H}'_0 \tag{35}$$

Using $\mathbf{H}_0$ and $\mathbf{H}'_0$ alone can rectify the image pair and achieve canonical epipolar geometry. However, combined with $\mathbf{H}_s$ and $\mathbf{H}'_s$ respectively, the appearance of the final rectification results can be improved further. A detailed description of the shearing transform can be found in [8], but its adaptation to our usage will be briefly described

below. We consider only one image, since the procedure can be carried to the other image identically.

For a given input image with width w and height h, coordinates of the midpoints on its four boundaries are shown in Fig. 4 and can be represented as $\mathbf{a}=\left[\frac{w-1}{2}\ 0\ 1\right]^T$, $\mathbf{b}=\left[w-1\ \frac{h-1}{2}\ 1\right]^T$, $\mathbf{c}=\left[\frac{w-1}{2}\ h-1\ 1\right]^T$, and $\mathbf{d}=\left[0\ \frac{h-1}{2}\ 1\right]^T$. The two central lines are expressed as $\mathbf{x}=\mathbf{b}-\mathbf{d}$, $\mathbf{y}=\mathbf{c}-\mathbf{a}$.

Let $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, $\hat{\mathbf{d}}$ be the coordinates of these four midpoints after transformation by $\mathbf{H}_0$, or

$$\hat{\mathbf{a}} = \mathbf{H}_0\mathbf{a} = \begin{bmatrix} \hat{a}_u \\ \hat{a}_v \\ \hat{a}_w \end{bmatrix} \xrightarrow{normalise} \begin{bmatrix} \hat{a}'_u \\ \hat{a}'_v \\ 1 \end{bmatrix},\ \hat{a}'_u = \frac{\hat{a}_u}{\hat{a}_w},\ \hat{a}'_v = \frac{\hat{a}_v}{\hat{a}_w} \tag{36}$$

The other three points ($\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, $\hat{\mathbf{d}}$) can be found in the same way, and the two central lines after rectification become $\hat{\mathbf{x}} = \hat{\mathbf{b}} - \hat{\mathbf{d}} = [x_u \quad x_v \quad 0]^T$, $\hat{\mathbf{y}} = \hat{\mathbf{c}} - \hat{\mathbf{a}} = [y_u \quad y_v \quad 0]^T$.

The perspective component of $\mathbf{H}_0$ causes the projective rectification to generate distortion, and the shearing transform $\mathbf{H}_s$, which is used to reduce the distortion, can be found by satisfying the following two constraints:

$$\text{Orthogonal}: \mathbf{x}^T\mathbf{y} = (\mathbf{H}_s\hat{\mathbf{x}})^T(\mathbf{H}_s\hat{\mathbf{y}}) = 0 \tag{37}$$

$$\text{Invariant aspect ratio}: \frac{\mathbf{x}^T\mathbf{x}}{\mathbf{y}^T\mathbf{y}} = \frac{(\mathbf{H}_s\hat{\mathbf{x}})^T(\mathbf{H}_s\hat{\mathbf{x}})}{(\mathbf{H}_s\hat{\mathbf{y}})^T(\mathbf{H}_s\hat{\mathbf{y}})} = \frac{\mathrm{w}^2}{\mathrm{h}^2} \tag{38}$$

where $\mathbf{H}_s$ is defined in (34). Expanding (37) and (38), and solving the quadratic polynomials of a and b based on the techniques similar to that were described in [8], we obtain

$$a = \frac{\mathrm{h}^2x_v^2 + \mathrm{w}^2y_v^2}{\mathrm{hw}(x_vy_u - x_uy_v)},\ b = \frac{\mathrm{h}^2x_ux_v + \mathrm{w}^2y_uy_v}{\mathrm{hw}(x_uy_v - x_vy_u)} \tag{39}$$

Substituting a and b into (34) yields the shearing transform matrix $\mathbf{H}_s$. Elements a' and b' of matrix $\mathbf{H}'_s$ can be found by the same way. After $\mathbf{H}_s$ and $\mathbf{H}'_s$ were obtained, the final homographies used for rectification with minimised geometrical distortion become $\mathbf{H} = \mathbf{H}_s\mathbf{H}_0$ and $\mathbf{H}' = \mathbf{H}'_s\mathbf{H}'_0$. They then can be used for resampling to complete the process of projective rectification.

3.4 Determination of image extent and resampling

In the projective transformation process, some of the image pixels might be converted to coordinate values that are outside the original image extent. In order to make these pixels visible, additional uniform scaling and translation can be applied [8], or the proper dimensions of the output image should be determined [7]. Although all the other rectification methods extend their image extent after the homographies computation and before the resampling, we propose a pre-extension scheme in our algorithm. This approach will extend the image pair by appending zero-value pixels around the boundary before determining the homographies ($\mathbf{H}$ and $\mathbf{H}'$). This modification, unexpectedly, can greatly reduce geometric distortion on some image pairs. If the size of the input image is $M \times N$, then the extended image size is set to $P \times P$, where $P \geqslant \sqrt{M^2 + N^2}$. These extra pixels are divided in half and append along the $\mathbf{x}$-axis (left, right) and $\mathbf{y}$-axis (top, bottom), as shown in Fig. 5.

Let (I, I') be the input image pair and be pre-extended to (I_e, I'_e). The output image pair after resampling is (J_e, J'_e). We will use the image I_e to describe the resampling process as follows. For each pixel location $\bar{\mathbf{m}}$ (with integer coordinates) in J_e, its value is set to the same as that of the pixel at position $\mathbf{m} = \mathbf{H}^{-1}\bar{\mathbf{m}}$ in I_e. Because point $\mathbf{m}$ usually does not have integer coordinates, its four neighbouring pixels are used to determine the value by linear interpolation. If $\mathbf{m}$ lies outside the extent of I_e, pixel $\bar{\mathbf{m}}$ is set to the chosen background colour.

3.5 Outline of algorithm

The rectification by using projective transformation will now be summarised. A pair of images, which contains a common overlap region is given as input. The output is a resampled pair of images, which satisfies the canonical epipolar geometry condition. The outline of the algorithm is as follows.

1. Extend the range of both input images.
2. Identify a set of point correspondences between the two size-extended images. At least six points are needed, though more can be used. Each point is found semi-automatically, i.e. it is chosen approximately by hand and a precise (subpixel accuracy) feature point close-by is then detected by the Harris corner finder [21] inside a local search window.
3. Determine the rectifying matrices ($\mathbf{H}_0$, $\mathbf{H}'_0$) and shearing transforms ($\mathbf{H}_s$, $\mathbf{H}'_s$) based on the algorithms described in Sections 3.3 and 3.4. The final projective transform matrices are $\mathbf{H} = \mathbf{H}_s\mathbf{H}_0$ and $\mathbf{H}' = \mathbf{H}'_s\mathbf{H}'_0$.
4. Resample the left and right extended images according to $\mathbf{H}$ and $\mathbf{H}'$, respectively. Extra black boundary pixels in the resampled images can be trimmed away.

4 Experimental results and discussion

To evaluate performance of the proposed method, various types of indoor and outdoor image pairs are tested. For each image pair, a set of 10 point correspondences were found semi-automatically from each image and used to compute two H matrices for 2-D projective rectification. To achieve the best results, these points are distributed evenly over the entire image area. The rectified results are shown in the following Section.

4.1 Visual evaluation of rectifying results

The original and rectified image pairs are shown in Figs. 6–9 for visual comparison. In order to make visual evaluation of the rectified results more convenient, four horizontal lines

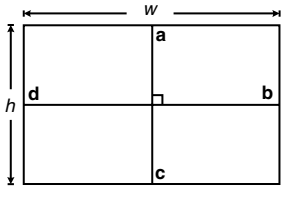

Fig. 4 *Centre points used for computing shearing transform matrices*

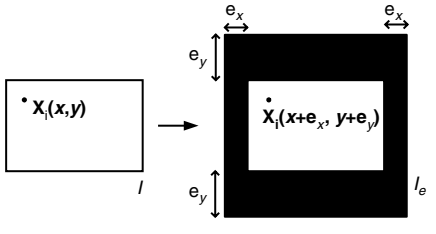

Fig. 5 *Image size pre-extension before homography computation and rectification*

are added to identify the differences before and after rectification.

4.1.1 Image pair 1: (outdoor) scene with far distance: The first pair used is aerial images acquired from the results published in [7]. These two images were taken on different days and are used to test our new algorithm for scenes of far distance. The rectified results are shown in Fig. 6, where the top two images are the original pair, the middle two images are rectified results using our method, and the bottom two images are rectified sections shown in Hartley's paper [7]. Apparently, the original image pair has been fully rectified by our method. Because there is almost no geometric distortion, these two images can be merged to create a 3-D impression. As mentioned in [7], when these two rectified images are viewed

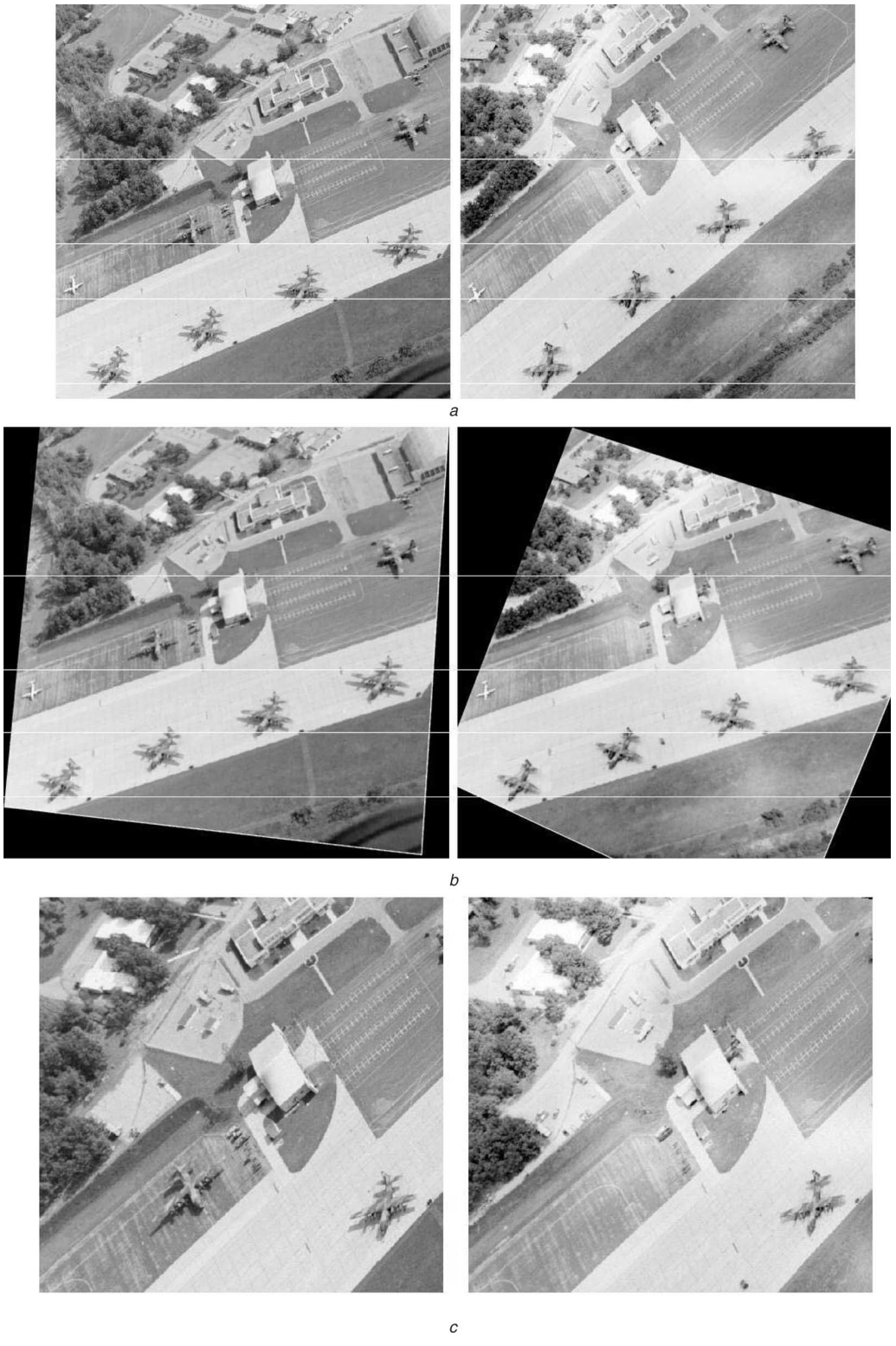

Fig. 6 *Rectifying results of airfield image pair used in [7]*

The white straight lines are used to mark the difference of y-coordinates before and after rectification
a Original pair of images
b Rectified images using the proposed method
c Sections of rectified images shown in Hartley's paper [7]

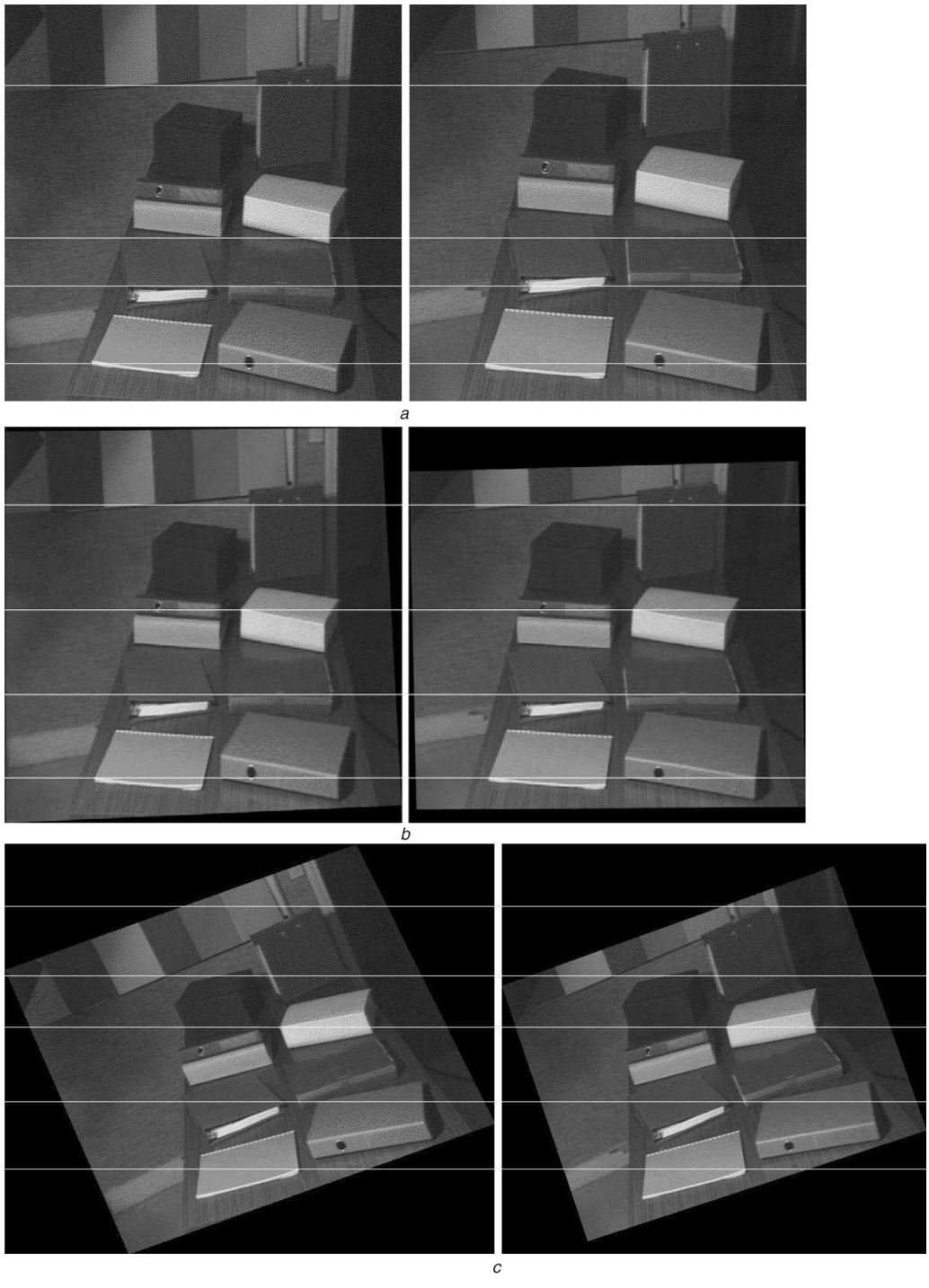

Fig. 7 *Rectifying results of the Books image pair*

a Original pair of images
b Rectified images using the proposed method
c Rectified images using IT method [13]

stereoscopically, the difference becomes apparent and may be used as an aid in detecting changes.

4.1.2 Image pairs 2 and 3: (indoor) scene with close distance: The second and third image pairs are indoor scene used in the IT method [13]. These two pairs were obtained from Isgrò's web site (see Acknowledgments) and are illustrated in the top rows of Figs. 7 and 8. The middle rows of these two Figures are the rectified images based on our method, and the bottom rows are the results rectified by the IT method. The comparisons show that our method does have less geometric distortion. While low geometric distortion may not be necessary for correspondence problems, it can greatly improve the stereoscopic viewing perception.

4.1.3 Image pair 4: (outdoor) scene with medium distance: The fourth image pair is an outdoor scene taken with a digital camera by ourselves. It is used to test the rectifying performance of an image pair shot at medium distance. While the image pair contains significant depth of field, the rectified result is satisfactory. It also shows that digital cameras can be used conveniently to take stereoscopic images if image rectification is available.

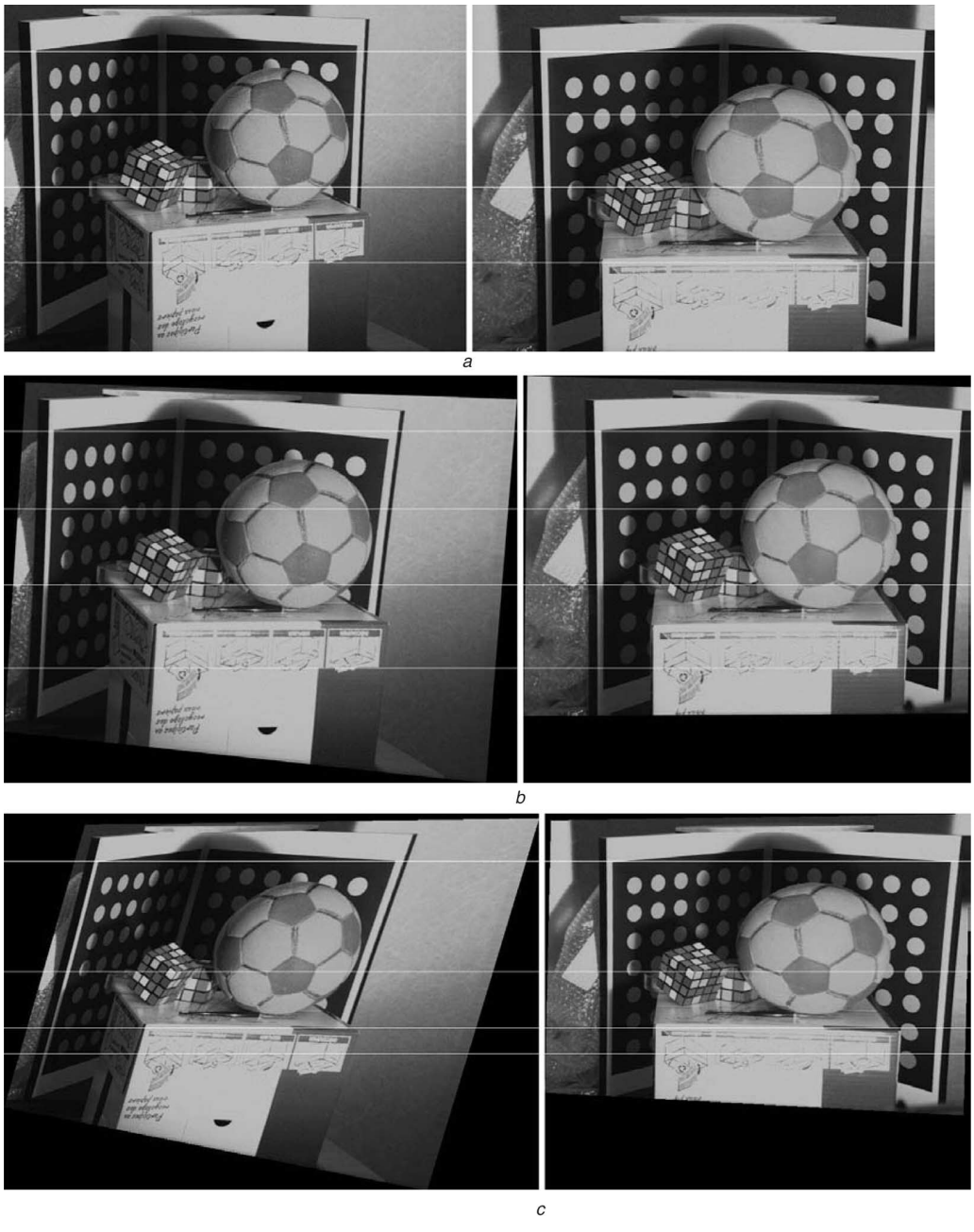

Fig. 8 *Rectifying results of the BalMouss image pair*

a Original pair of images
b Rectified images using the proposed method
c Rectified images using IT method [13]

4.2 Quantitative evaluation of rectifying results and time complexity analysis

In addition to the above visual comparisons, quantitative evaluation based on the changes of y disparity is also conducted. To estimate the accuracy of the rectification process numerically, the value of the criterion described in (19) or $\bar{\mathbf{m}}'^T\mathbf{F}_\infty\bar{\mathbf{m}} = (\mathbf{H}'\mathbf{m}')^T\mathbf{F}_\infty(\mathbf{H}\mathbf{m})$ is used. This represents the y disparity between a given point correspondence pair, $(\mathbf{m}, \mathbf{m}')$, after rectification. Defining the mean of the absolute difference (MAD) of y coordinate, $|\overline{\Delta y}|$, for the original and rectified image pairs as

$$|\overline{\Delta y}| = \begin{cases} \varepsilon_{org} = \frac{1}{N}\sum_{i=1}^{N} |(\mathbf{m}_i)_y - (\mathbf{m}'_i)_y| & : \quad \text{original} \\ \varepsilon_{rec} = \frac{1}{N}\sum_{i=1}^{N} |(\mathbf{H}\mathbf{m}_i)_y - (\mathbf{H}'\mathbf{m}'_i)_y| & : \quad \text{rectified} \end{cases}$$

where $(\cdot)_y$ indicates the y coordinates and $(\mathbf{m}_i, \mathbf{m}'_i)$ represents the coordinates of the ith pair of the point correspondences. A desirable property of the rectification is to minimise the value of $|\overline{\Delta y}|$. When the ten chosen point correspondences, that have been used in computation of homographies (training data), were used for $|\overline{\Delta y}|$ computation, the results before and after rectification for the above four image pairs are listed in Table 1. If an image pair is ideally rectified, $|\overline{\Delta y}|$ of the point correspondences after rectification should be zero. As can be seen from the Table, for all the training image pairs, values of $|\overline{\Delta y}|$ after rectification of are all less than 1 pixel. They are greatly reduced compared with the values before rectification (ε_{org}), indicating the effectiveness of the proposed method. To further verify the rectifying results, ten other randomly chosen point correspondences (testing data) not included in the original training set are assessed for the values of $|\overline{\Delta y}|$,

Fig. 9 *Rectifying results of outdoor scene taken by digital camera at a medium distance*

and the results are also summarised in Table 1. As can be observed from the Table, $|\overline{\Delta y}|$ will be always less than one pixel for the pairs in the training set and around one pixel for the pairs in the testing set.

The time complexity of the proposed algorithm involves two parts. The first part is the computation of the homographies, and the second part is the resampling. The solution for the pair of homogphies with ten point correspondences used for optimisation can be found in less than six iterations by the Levenberg-Marquardt algorithm. The iteration stops if the mean square error (MSE), $\frac{1}{N}\sum_{i=1}^{N} w_i \frac{\left(\mathbf{m}_i'^T \mathbf{F} \mathbf{m}_i\right)^2}{2}$, or $\frac{1}{2N}\left\|\mathbf{W}^{1/2}\mathbf{U}_N^T\mathbf{f}\right\|^2$ satisfies at least one of the following two conditions (1) MSE is less than 10^{-3}, (2) the difference between two successive MSE is less than 10^{-5}. Using a Pentium-4, 2.8 GHz PC, the time of this iteration takes less than 0.1 s. Note that if the configuration of the camera pair remains unchanged, the homographies could stay the same for all the image pairs taken.

The actual computation time of resampling for each image pair, which depends on the original size of the input image, is shown in Table 2. Since this process is pretty standard, the time costs for different methods would be quite similar. In order to avoid clipping during resampling, the image is extended as has been described in Fig. 5. This extension of size unavoidably increases the time of computation. The actual size and the resampling time for the extended images are also shown in Table 2.

Table 1: MAD of *y*-coordinate before (ϵ_{org}) and after (ϵ_{rec}) rectification with the proposed method, where the average of $|\Delta y|$ is based on 10 point correspondences. The evaluation is performed on both training data set and testing data set

Image name	Pair 1	Pair 2	Pair 3	Pair 4
$\|\overline{\Delta y}\|$	Airfield	Books	Balmouss	Outdoor
ε_{org}(pixel)	16.10	2.4	10.20	23.9
ε_{rec}(pixel) training data	0.37	0.21	0.72	0.38
ε_{rec}(pixel) testing data	0.60	0.93	0.98	0.91

4.3 Comparisons and discussion

Compared with the results presented in Isgrò and Trucco [13] (IT method) and Hartley [7], our new method has the following advantages:

(1) Because our algorithm requires fewer parameters to be estimated in the nonlinear iterative-minimisation process, it needs fewer point correspondences for computation of homographies. The reduction of the number of parameters makes our algorithm more robust in rectification. Although we pick 10 matched points for all the image pairs shown in this paper to test our method, $6 \sim 8$ points are usually enough to obtain satisfactory results. Other methods use more points for each image pair to achieve the same result (16 matched points for Fig. 7 in [13] and 25 matched points for Fig. 6 in [7]).

(2) In order to obtain a unique solution, the IT method uses (14) to minimise the **x**-axis disparity. This extra constraint will be undesirable if the rectified result is used for stereoscopic viewing. Instead of minimising **x**-axis disparity, a shearing transform is used in our algorithm to preserve aspect ratio and perpendicularity of the central axes. It helps to reduce geometric distortion and improve viewing sensation.

(3) Initial values of the nonlinear solution by iteration are much easier to set. No initial solution needs to be estimated, that is, the initial parameter vector $\mathbf{\Phi}_0$ is simply set to $[1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0]^T$.

Table 2: Actual computation time for the image pairs under test. The time cost is dependent on the size of the input image. The image is usually extended to avoid clipping during resampling

Name and size Time (s)	Pair 1 Airfield	Pair 2 Books	Pair 3 Balmouss	Pair 4 Outdoor
Original size	0.27 (430 × 430)	0.39 (512 × 512)	0.634 (576 × 768)	0.434 (480 × 640)
Extended size	0.56(610 × 610)	0.84 (725 × 725)	1.33(960 × 960)	0.94 (800 × 800)

Most of the projective rectification methods proposed all base their algorithms on an estimation of the F matrix. However, different estimators used will obtain different F matrices, which leads to varying rectification results. Our approach, similar to the IT method [13], avoids an F matrix estimation procedure and obtains homographies directly and uniquely. Furthermore, it improves on the IT method and obtains better rectified results with fewer matched points and lower geometric distortion.

5 Conclusion

By incorporating the idea of relative modification, the proposed method parameterises homography with fewer parameters to be estimated. Formulating purposely to obtain the H matrices directly, we avoid the necessity of camera calibration and computation of an F matrix. Furthermore, instead of putting constraint on **x**-axis disparity, we use a shearing transform to achieve a unique solution for the projective rectification problem, and greatly reduce the geometric distortion. The effectiveness of our method has been verified by an extensive set of real experiments, and four sets of image pairs are demonstrated in the paper. Visual inspection confirms the accuracies and low geometric distortion of the proposed method, and quantitative results show that the y disparity after rectification can be reduced to less than one pixel. These four sets of image pairs represent different types of imaging geometry and scene range, and the results demonstrate that the proposed method can achieve satisfactory results with acceptable time complexity. Comparisons with two other methods are also provided to illustrate its performance.

6 Acknowledgments

The authors gratefully acknowledge the helpful suggestions made by the anonymous reviewers for improving the paper from its earlier draft. They also wish to express their appreciation for the support of the National Science Council under NSC 92-2213-E-224-014.

Stereo image pairs in Figs. 7 and 8 are kindly available from the following web site under copyright http://www.math.unipa.it/ ~ fisgro/research/demorectification/demorectification.html

7 References

1 Faugeras, O.: 'Three-dimensional computer vision: A geometric viewpoint' (MIT Press, Cambridge, MA, USA, 1993)
2 Slama, C.: 'Manual of photogrammetry' (American Society of Photogrammetry, 1980)
3 Ayache, N., and Hansen, C.: 'Rectification of images for binocular and trinocular stereovision'. IEEE ICPR, Rome, Italy, 1988, pp. 11–16
4 Ayache, N., and Lustman, F.: 'Trinocular stereo vision for robotics', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991, **13**, pp. 73–85
5 Fusiello, A., Trucco, E., and Verri, A.: 'A compact algorithm for rectification of stereo pairs', *Mach. Vis. Appl.*, 2000, **12**, pp. 16–22
6 Robert, L., Zeller, C., Faugeras, O., and Hebert, M.: 'Applications of non-metric vision to some visually-guided robotics tasks', in Aloimonos, Y., (Ed.): 'Visual navigation: From biological systems to unmanned ground vehicles' (Advances in Computer Vision, Lawrence Erlbaum Associates, 1997) Vol. II
7 Hartley, R.: 'Theory and practice of projective rectification', *Int. J. Comput. Vis.*, 1999, **35**, pp. 115–127
8 Loop, C., and Zhang, Z.: 'Computing rectifying homographies for stereo vision'. IEEE CVPR', Fort Collins, CO, USA, 1999, pp. 125–131
9 Pollefeys, M., Kock, R., and Gool, L.V.: 'A simple and efficient rectification method for general motion'. Proc. ICCV, Corfu, Greece, 1999, pp. 496–501
10 Chen, Z., Wu, C., and Tsui, H.T.: 'A new image rectification algorithm', *Pattern Recognit. Lett.*, 2003, **24**, (1-3), pp. 251–260
11 Al-Shalfan, K.A., Haigh, J.G.B., and Ipson, S.S.: 'Direct algorithm for rectifying pairs of uncalibrated images', *Electron. Lett.*, 2000, **36**, (5), pp. 419–420
12 Papadimitriou, D.V., and Dennis, T.J.: 'Epipolar line estimation and rectification for stereo image pairs', *IEEE Trans. Image Process.*, 1996, **5**, (4), pp. 672–676
13 Isgro, F., and Trucco, E.: 'Projective rectification without epipolar geometry'. Proc. IEEE CVPR, Fort Collins (Colorado), USA, 1999, **I**, pp. 94–99
14 Hartley, R.: 'In defense of the eight-point algorithm', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 580–593
15 Zhang, Z.: 'Determining the epipolar geometry and its uncertainty A review', *Int. J. Comput. Vis.*, 1998, **27**, pp. 161–195
16 Hartley, R., and Zisserman, A.: 'Multiple view geometry in computer vision' (Cambridge University Press, 2000)
17 Dimension Technologies, http://www.dti3d.com/
18 StereoGraphics corporation, http://www.stereographics.com
19 Lei, B., Chang, C., and Hendriks, E. A.: 'An efficient image-based telepresence system for videoconferencing', *IEEE Trans. Circuit Syst. Video Technol.*, 2004, **14**, (3), pp. 335–348
20 Lei, B., and Hendriks, E.: 'A real-time realization of geometrical valid view synthesis for tele-conferencing with viewpoint adaptation, 3D data, processing'. Visualization and Transmission Conf., Padavo, Italy, 2002, pp. 327–336
21 Harris, C., and Stephens, M.: 'A combined corner and edge detector'. Alvey Vision Conf., 1998, pp. 147–151