ELSEVIER

Artificial Intelligence 78 (1995) 87-119

Artificial Intelligence

A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry

Zhengyou Zhang *, Rachid Deriche, Olivier Faugeras, Quang-Tuan Luong

INRIA Sophia-Antipolis, 2004 route des Lucioles, B.P. 93, F-06902 Sophia-Antipolis Cedex, France

Received June 1994; revised December 1994

Abstract

This paper proposes a robust approach to image matching by exploiting the only available geometric constraint, namely, the epipolar constraint. The images are uncalibrated, namely the motion between them and the camera parameters are not known. Thus, the images can be taken by different cameras or a single camera at different time instants. If we make an exhaustive search for the epipolar geometry, the complexity is prohibitively high. The idea underlying our approach is to use classical techniques (correlation and relaxation methods in our particular implementation) to find an initial set of matches, and then use a robust technique—the Least Median of Squares (LMedS)—to discard false matches in this set. The epipolar geometry can then be accurately estimated using a meaningful image criterion. More matches are eventually found, as in stereo matching, by using the recovered epipolar geometry. A large number of experiments have been carried out, and very good results have been obtained.

Regarding the relaxation technique, we define a new measure of matching support, which allows a higher tolerance to deformation with respect to rigid transformations in the image plane and a smaller contribution for distant matches than for nearby ones. A new strategy for updating matches is developed, which only selects those matches having both high matching support and low matching ambiguity. The update strategy is different from the classical "winner-take-all", which is easily stuck at a local minimum, and also from "loser-take-nothing", which is usually very slow. The proposed algorithm has been widely tested and works remarkably well in a scene with many repetitive patterns.

Keywords: Robust matching; Epipolar geometry; Fundamental matrix; Least Median of Squares (LMedS); Relaxation; Correlation

* Corresponding author. E-mail: zzhang@@sophia.inria.fr.

SSDI 0004-3702(95)00022-4

1. Introduction

Matching different images of a single scene remains one of the bottlenecks in computer vision. A large amount of work has been carried out during the last 15 years, but the results are, however, not satisfactory. The only geometric constraint we know between two images of a single scene is the *epipolar constraint*. However, when the motion between the two images is not known, the epipolar geometry is also unknown. The methods reported in the literature all exploit some heuristics in one form or another, for example, intensity similarity, which are not applicable to most cases. The difficulty is partly bypassed by taking long sequences of images over a short time interval [10, 57]. Indeed, as the time interval is small and object velocity is constrained by physical laws, the interframe displacements of objects are bounded, i.e., the correspondence of a token at the subsequent instant must be in the neighborhood of the first. However, in many cases, such as a pair of uncalibrated stereo images, this technique does not apply. Developing a robust image matching technique is thus very important.

Over the years numerous algorithms for image matching have been proposed. They can roughly be classified into two categories:

(1) *Template matching.* In this category, the algorithms attempt to correlate the grey levels of image patches in the views being considered, assuming that they present some similarity [4, 7, 14–16]. The underlying assumption appears to be a valid one for relatively textured areas and for image pairs with small difference; however it may be wrong at occlusion boundaries and within featureless regions.

(2) *Feature matching.* In this category, the algorithms first extract salient primitives from the images, such as edge segments or contours, and match them in two or more views. An image can then be described by a graph with primitives defining the nodes and geometric relations defining the links. The registration of two maps becomes the mapping of the two graphs: *subgraph isomorphism.* Common techniques are tree searching, relaxation, maximal clique detection, etc. Some heuristics such as assuming affine transformation between the two images are usually introduced to reduce the complexity. These methods are fast because only a small subset of the image pixels are used, but may fail if the chosen primitives cannot be reliably detected in the images. The following list of references is by no means exhaustive: [5, 6, 22, 35, 45, 50, 54]

The approach we propose in this paper aims at exploiting the only geometric constraint, i.e., the epipolar constraint, to establish robustly correspondences between two perspective images of a single scene. However, in order to reduce the complexity of the algorithm, we still exploit heuristic techniques to find an initial set of matches. We first extract high curvature points and then match them using a classical correlation technique followed by a new fuzzy relaxation procedure. More precisely, our algorithm consists of three steps:

- establish initial correspondences using some classical techniques,
- estimate robustly the epipolar geometry,
- establish correspondences using estimated epipolar geometry as in stereo matching.

The basic idea is first to estimate robustly the epipolar geometry, and then reduce the general image matching problem to stereo matching. In the subsequent sections, we will

first review the epipolar geometry, and then describe in detail the three steps of the proposed approach. A preliminary version of this paper appeared in the Proceedings of the Third European Conference on Computer Vision [12].

A similar idea has been independently exploited by Xu et al. [40,56], who also searched for image correspondences through the recovery of the epipolar geometry. There are however two main differences:

- The weak perspective camera model is used in their work, and a full perspective model is used in ours. The choice of the most appropriate criterion for the recovery of the epipolar geometry is not addressed in their work.
- The search for the epipolar geometry is carried out with an exhaustive strategy in their work. The complexity is prohibitively high even for a weak perspective model ($O(m^4n^4)$, where m and n are the number of points in the first and second image, respectively). The complexity is reduced by checking only a few closest points. In our work, some classical techniques are applied to find an initial set of correspondences.

We could apply the same strategy as that of Xu et al. [40,56]. In fact, it has been applied to solve the correspondence problem between two sets of 3D line segments [58]. When applying it to the problem addressed in this paper, we need 8 point correspondences in order to estimate the epipolar geometry, because we use a full perspective model. The complexity is then $O(m^8n^8)$. Suppose both m and n are 100, the complexity is in the order of 10^{32}! Xu et al. [40,56] deal with also the motion segmentation problem using epipolar constraint, which is not addressed in this paper.

Recently, computer vision researchers have paid much attention to the robustness of vision algorithms because the data are unavoidably error prone [17,59]. Many the so-called *robust regression* methods have been proposed that are not so easily affected by outliers [25,48]. The reader is referred to [48, Chapter 1] for a review of different robust methods. The two most popular robust methods are the *M-estimators* and the *Least Median of Squares* (LMedS) method (see Section 6.3). Kumar and Hanson [26] compared different robust methods for pose refinement from 3D–2D line correspondences, while Meer et al. [38], for image smoothing. Haralick et al. [18] applied M-estimators to solve the pose problem from point correspondences. Thompson et al. [51] applied the LMedS estimator to detect moving objects using point correspondences between orthographic views. Other recent works on the application of robust techniques to motion segmentation include [3,42,52].

Regarding the robust recovery of the epipolar geometry, our work is closely related to that of Olsen [43] and that of Shapiro and Brady [49]. Olsen uses a linear method to estimate the epipolar geometry, which has already been shown in one of our previous work [32] to be insufficiently accurate. He further assumes that knowledge of the epipolar geometry, as in many practical cases, is available. In particular, he assumes the epipolar lines are almost aligned horizontally. This knowledge is then used to find matches between the stereo image pair, and a robust method (the M-estimator, see Section 6.3) is used to detect false matches and to obtain a better estimate of the epipolar geometry. Shapiro and Brady also use a linear method. The camera model is however a simplified one, namely an affine camera. Correspondences are established by tracking corner features over time. False matches are rejected through a *regression diagnostic*,

which computes an initial estimate of the epipolar geometry over all matches, and sees how the estimate changes if a match is deleted. The match whose removal maximally reduces the residual is identified to be an *outlier* and is rejected. The procedure is then repeated with the reduced set of matches until all outliers have been removed. These two approaches (M-estimators and regression diagnostics) work well when the percentage of outliers is small and more importantly when their derivations from the valid matches are not too large, as in the above two works. In the case described in this paper, two images can be quite different. There may be a large percentage of false matches (usually around 20%, sometimes 40%) using heuristic matching techniques such as correlation, and a false match may be completely different from the valid matches. The robust technique described in this paper deals with these issues and can theoretically detect outliers when they make up as much as 50% of whole data.

2. Notation

A camera is described by the widely used pinhole model. The coordinates of a 3D point $M = [x, y, z]^T$ in a world coordinate system and its retinal image coordinates $\boldsymbol{m} = [u, v]^T$ are related by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbb{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

where s is an arbitrary scale, and $\mathbb{P}$ is a 3×4 matrix, called the perspective projection matrix. Denoting the homogeneous coordinates of a vector $\boldsymbol{x} = [x, y, \ldots]^T$ by $\tilde{\boldsymbol{x}}$, i.e., $\tilde{\boldsymbol{x}} = [x, y, \ldots, 1]^T$, we have $s\tilde{\boldsymbol{m}} = \mathbb{P}\tilde{M}$.

The matrix $\mathbb{P}$ can be decomposed as

$$\mathbb{P} = \boldsymbol{A}[\boldsymbol{R}\ \boldsymbol{t}],$$

where $\boldsymbol{A}$ is a 3×3 matrix, mapping the normalized image coordinates to the retinal image coordinates, and $(\boldsymbol{R}, \boldsymbol{t})$ is the 3D displacement (rotation and translation) from the world coordinate system to the camera coordinate system. The most general matrix $\boldsymbol{A}$ can be written as

$$\boldsymbol{A} = \begin{bmatrix} -fk_u & fk_u \cot\theta & u_0 \\ 0 & -fk_v/\sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where

- f is the focal length of the camera,
- k_u and k_v are the horizontal and vertical scale factors, whose inverses characterize the size of the pixel in the world coordinate unit,
- u_0 and v_0 are the coordinates of the principal point of the camera, i.e., the intersection between the optical axis and the image plane, and

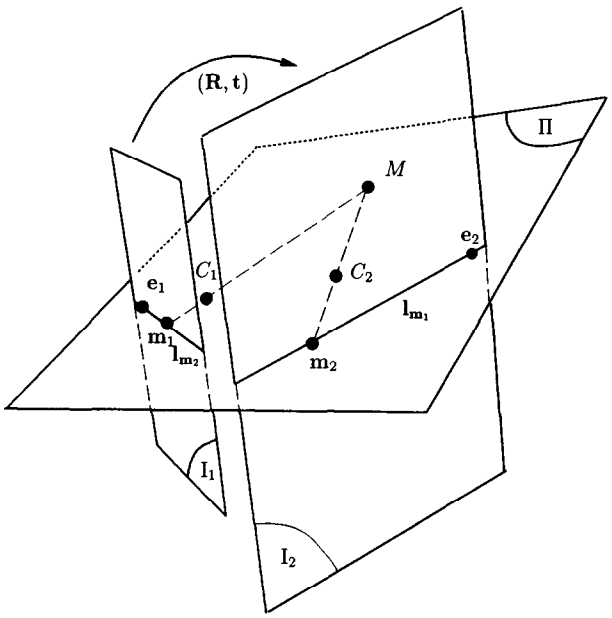

Fig. 1. The epipolar geometry.

- θ is the angle between the retinal axes. (This parameter is introduced to account for the fact that the pixel grid may not be exactly orthogonal. In practice, however, it is very close to $\pi/2$.)

As is clear, we cannot separate f from k_u and k_v. We thus have five intrinsic parameters for each camera: $\alpha_u = -fk_u$, $\alpha_v = -fk_v$, u_0, v_0 and θ.

The first and second images are respectively denoted by I_1 and I_2. A point $\boldsymbol{m}$ in the image plane I_i is noted as $\boldsymbol{m}_i$. The second subscript, if any, will indicate the index of the point in consideration.

3. Epipolar geometry

Considering the case of two cameras as shown in Fig. 1.

Let C_1 and C_2 be the optical centers of the first and second cameras, respectively. Given a point $\boldsymbol{m}_1$ in the first image, its corresponding point in the second image is constrained to lie on a line called the *epipolar line* of $\boldsymbol{m}_1$, denoted by $\boldsymbol{l}_{\boldsymbol{m}_1}$. The line $\boldsymbol{l}_{\boldsymbol{m}_1}$ is the intersection of the plane Π, defined by $\boldsymbol{m}_1$, C_1 and C_2 (known as the *epipolar plane*), with the second image plane I_2. This is because image point $\boldsymbol{m}_1$ may correspond to an arbitrary point on the semi-line C_1M (M may be at infinity) and that the projection of C_1M on I_2 is the line $\boldsymbol{l}_{\boldsymbol{m}_1}$. Furthermore, one observes that all epipolar lines of the points in the first image pass through a common point $\boldsymbol{e}_2$, which is called the *epipole*. $\boldsymbol{e}_2$ is the intersection of the line C_1C_2 with the image plane I_2. This can be easily understood as follows. For each point $\boldsymbol{m}_{1k}$ in the first image I_1, its epipolar line $\boldsymbol{l}_{\boldsymbol{m}_{1k}}$ in

I_2 is the intersection of the plane Π^k, defined by $\boldsymbol{m}_{1k}$, C_1 and C_2, with image plane I_2. All epipolar planes Π^k thus form a pencil of planes containing the line C_1C_2. They must intersect I_2 at a common point, which is $\boldsymbol{e}_2$. Finally, one can easily see the symmetry of the epipolar geometry. The corresponding point in the first image of each point $\boldsymbol{m}_{2k}$ lying on $\boldsymbol{l}_{m_{1k}}$ must lie on the epipolar line $\boldsymbol{l}_{m_{2k}}$, which is the intersection of the same plane Π^k with the first image plane I_1. All epipolar lines form a pencil containing the epipole $\boldsymbol{e}_1$, which is the intersection of the line C_1C_2 with the image plane I_1. The symmetry leads to the following observation. If $\boldsymbol{m}_1$ (a point in I_1) and $\boldsymbol{m}_2$ (a point in I_2) correspond to a single physical point M in space, then $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, C_1 and C_2 must lie in a single plane. This is the well-known *co-planarity constraint* or *epipolar equation* in solving motion and structure from motion problems when the intrinsic parameters of the cameras are known [29].

Let the displacement from the first camera to the second be $(\boldsymbol{R}, \boldsymbol{t})$. Let $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ be the images of a 3D point M on the cameras. Without loss of generality, we assume that M is expressed in the coordinate frame of the first camera. Under the pinhole model, we have the following two equations:

$$s_1\tilde{\boldsymbol{m}}_1 = \boldsymbol{A}_1 [\boldsymbol{I}\ \boldsymbol{0}] \begin{bmatrix} M \\ 1 \end{bmatrix}, \qquad s_2\tilde{\boldsymbol{m}}_2 = \boldsymbol{A}_2 [\boldsymbol{R}\ \boldsymbol{t}] \begin{bmatrix} M \\ 1 \end{bmatrix},$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are the intrinsic matrices of the first and second cameras, respectively. Eliminating M, s_1 and s_2 from the above equations, we obtain the following fundamental equation

$$\tilde{\boldsymbol{m}}_2^{\mathrm{T}} \boldsymbol{A}_2^{-\mathrm{T}} \boldsymbol{T} \boldsymbol{R} \boldsymbol{A}_1^{-1} \tilde{\boldsymbol{m}}_1 = 0, \tag{2}$$

where $\boldsymbol{T}$ is an antisymmetric matrix defined by $\boldsymbol{t}$ such that $\boldsymbol{T}\boldsymbol{x} = \boldsymbol{t} \wedge \boldsymbol{x}$ for all 3D vectors $\boldsymbol{x}$ ($\wedge$ denotes the cross product).

Eq. (2) is a fundamental constraint underlying any two images if they are perspective projections of one and the same scene. Let $\boldsymbol{F} = \boldsymbol{A}_2^{-\mathrm{T}} \boldsymbol{T} \boldsymbol{R} \boldsymbol{A}_1^{-1}$, $\boldsymbol{F}$ is known as the fundamental matrix of the two images [31]. Without considering 3D metric entities, we can think of the fundamental matrix as providing the two epipoles (i.e., the vertexes of the two pencils of epipolar lines) and the three parameters of the homography between these two pencils. This is the only geometric information available from two uncalibrated images [31,36]. This implies that the fundamental matrix has only seven degrees of freedom. Indeed, it is only defined up to a scale factor and its determinant is zero. Geometrically, $\boldsymbol{F}\tilde{\boldsymbol{m}}_1$ defines the epipolar line of point $\boldsymbol{m}_1$ in the second image. Eq. (2) says no more than that the correspondence in the right image of point $\boldsymbol{m}_1$ lies on the corresponding epipolar line. Transposing Eq. 2 yields the symmetric relation from the second image to the first image.

It can be shown that the fundamental matrix $\boldsymbol{F}$ is related to the essential matrix $\boldsymbol{E} = \boldsymbol{t} \times \boldsymbol{R}$ [23,29] by

$$\boldsymbol{F} = \boldsymbol{A}_2^{-\mathrm{T}} \boldsymbol{E} \boldsymbol{A}_1^{-1}.$$

It is thus clear that if the cameras are calibrated, the problem becomes the one of *motion and structure from motion* [1,13,24,29,39,53,55].

4. Finding candidate matches by correlation

Before recovering the epipolar geometry, we must establish a few correspondences between images. To this end, a corner detector is first applied to each image to extract high curvature points. A classical correlation technique is then used to establish matching candidates between the two images.

4.1. Extracting points of interest

First, feature points corresponding to high curvature points are extracted from each image. A great deal of effort has been spent by the computer vision community on this problem, and several approaches have been reported in the literature in the last few years. They can be broadly divided into two groups: The first group consists in first extracting edges as a chain code, and then searching for points having maxima curvature [2,9,37] or performing a polygonal approximation on the chains and then searching for the line segment intersections [21]. The second group works directly on a grey-level image. The large number of techniques that have been proposed within this group are generally based on the measurement of the gradients and of the curvatures of the surface (see [11] for a review).

In our application, we use the corner detector [20], which is a slightly modified version of the Plessey corner detector [19,41]. It is based on the following operator:

$$R(x,y) = \det[\hat{C}] - k\,\mathrm{trace}^2[\hat{C}], \tag{3}$$

where $\hat{C}$ is the following matrix:

$$\hat{C} = \begin{bmatrix} \hat{I}_x^2 & \widehat{I_x I_y} \\ \widehat{I_x I_y} & \hat{I}_y^2 \end{bmatrix}, \tag{4}$$

where $\hat{I}$ denotes the smoothing operation on the grey-level image $I(x,y)$. I_x and I_y indicate the x and y directional derivatives respectively.

We use a value of k equal to 0.04 to provide discrimination against high contrast pixel step edges. After that, the operator output is thresholded for the corner detection. It should be pointed out that this method allows us to recover a corner position up to pixel precision. In order to recover the corner position up to subpixel position, one uses the model based approach we have already developed and presented in [8], where corners are extracted directly from the image by searching the parameters of the parametric model that best approximate the observed grey-level image intensities around the corner position detected. One can notice that such an approach is well adapted for scenes containing polyhedral objects, but not for most outdoor scenes.

4.2. Matching through correlation

Given a high curvature point $\boldsymbol{m}_1$ in image 1, we use a correlation window of size $(2n+1) \times (2m+1)$ centered at this point. We then select a rectangular search area of size $(2d_u+1) \times (2d_v+1)$ around this point in the second image, and perform a

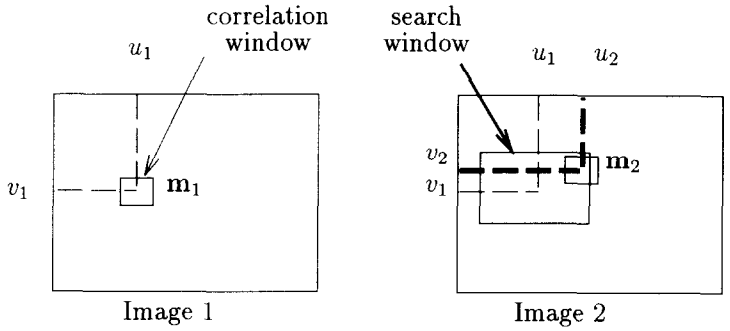

Fig. 2. Correlation.

correlation operation on a given window between point $\boldsymbol{m}_1$ in the first image and all high curvature points $\boldsymbol{m}_2$ lying within the search area in the second image. The search window reflects some *a priori* knowledge about the disparities between the matched points. This is equivalent to reducing the search area for a corresponding point from the whole image to a given window. The correlation score is defined as

$$Score(\boldsymbol{m}_1, \boldsymbol{m}_2) = \sum_{i=-n}^{n} \sum_{j=-m}^{m} \left([I_1(u_1 + i, v_1 + j) - \overline{I_1(u_1, v_1)}] \times [I_2(u_2 + i, v_2 + j) - \overline{I_2(u_2, v_2)}] \right) \times \left((2n+1)(2m+1)\sqrt{\sigma^2(I_1) \times \sigma^2(I_2)} \right)^{-1}, \tag{5}$$

where

$$\overline{I_k(u,v)} = \frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} I_k(u+i, v+j)}{(2n+1)(2m+1)}$$

is the average at point (u, v) of I_k $(k = 1, 2)$, and $\sigma(I_k)$ is the standard deviation of the image I_k in the neighborhood $(2n+1) \times (2m+1)$ of (u, v), which is given by:

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=-n}^{n} \sum_{j=-m}^{m} I_k^2(u,v)}{(2n+1)(2m+1)} - \overline{I_k(u,v)}}. \tag{6}$$

The score ranges from -1, for two correlation windows which are not similar at all, to 1, for two correlation windows which are identical.

A constraint on the correlation score is then applied in order to select the most consistent matches: For a given couple of points to be considered as a candidate match, the correlation score must be higher than a given threshold. If the above constraint is fulfilled, we say that the pair of points considered is self consistent and forms a *candidate match*. For each point in the first image, we thus have a set of candidate matches from the second image (the set is possibly nil); and in the same time we have also a set of candidate matches from the first image for each point in the second image.

In our implementation, $n = m = 7$ for the correlation window, and a threshold of 0.8 on the correlation score is used. For the search window, d_u and d_v are set to a quarter

of the image width and height, respectively. It is thus very large (quarter of the whole image).

5. Disambiguating matches through relaxation

Using the correlation technique described above, a point in the first image may be paired to several points in the second image (which we call *candidate matches*), and vice versa. Several techniques exist for resolving the matching ambiguities. The technique we use falls into the class of techniques known as *relaxation techniques*. The idea is to allow the candidate matches to reorganize themselves by propagating some constraints, such as continuity and uniqueness, through the neighborhood.

5.1. Measure of the support for a candidate match

Consider a candidate match $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$ where $\boldsymbol{m}_{1i}$ is a point in the first image and $\boldsymbol{m}_{2j}$ is a point in the second image. Let $\mathcal{N}(\boldsymbol{m}_{1i})$ and $\mathcal{N}(\boldsymbol{m}_{2j})$ be, respectively, the neighbors of $\boldsymbol{m}_{1i}$ and $\boldsymbol{m}_{2j}$ within a disc of radius R. If $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$ is a good match, we will expect to see many matches $(\boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})$, where $\boldsymbol{n}_{1k} \in \mathcal{N}(\boldsymbol{m}_{1i})$ and $\boldsymbol{n}_{2l} \in \mathcal{N}(\boldsymbol{m}_{2j})$, such that the position of $\boldsymbol{n}_{1k}$ relative to $\boldsymbol{m}_{1i}$ is similar to that of $\boldsymbol{n}_{2l}$ relative to $\boldsymbol{m}_{2j}$. On the other hand, if $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$ is a bad match, we will expect to see only few matches, or even not any at all, in their neighborhood.

More formally, we define a measure of support for a match, which we call the *strength of the match* (SM for abbreviation), as

$$S_M(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j}) = c_{ij} \sum_{\boldsymbol{n}_{1k} \in \mathcal{N}(\boldsymbol{m}_{1i})} \left[\max_{\boldsymbol{n}_{2l} \in \mathcal{N}(\boldsymbol{m}_{2j})} \frac{c_{kl}\, \delta(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})}{1 + dist(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})} \right],$$

where c_{ij} and c_{kl} are the goodness of the candidate matches $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$ and $(\boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})$, which can be the correlation scores given in the last section, $dist(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})$ is the average distance of the two pairings, i.e.,

$$dist(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l}) = [d(\boldsymbol{m}_{1i}, \boldsymbol{n}_{1k}) + d(\boldsymbol{m}_{2j}, \boldsymbol{n}_{2l})]/2$$

with $d(\boldsymbol{m}, \boldsymbol{n}) = \|\boldsymbol{m} - \boldsymbol{n}\|$, the Euclidean distance between $\boldsymbol{m}$ and $\boldsymbol{n}$, and

$$\delta(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l}) = \begin{cases} e^{-r/\varepsilon_r}, & \text{if } (\boldsymbol{n}_{1k}, \boldsymbol{n}_{2l}) \text{ is a candidate match and } r < \varepsilon_r, \\ 0, & \text{otherwise,} \end{cases}$$

where r is the relative distance difference given by

$$r = \frac{|d(\boldsymbol{m}_{1i}, \boldsymbol{n}_{1k}) - d(\boldsymbol{m}_{2j}, \boldsymbol{n}_{2l})|}{dist(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j};\ \boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})}$$

and ε_r is a threshold on the relative distance difference. The above definition of the strength of a match is similar in the form to that used in the PMF stereo algorithm [44].

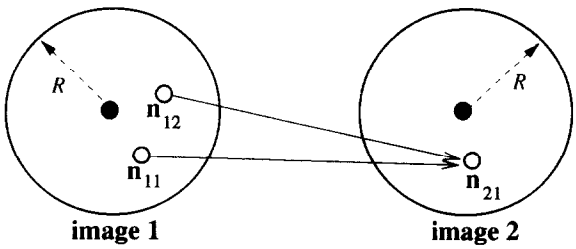

Fig. 3. Illustration of the non-symmetric problem of the matching support measure.

Several remarks can be made regarding our measure of matching support.

- First, the strength of a match actually counts the number of candidate matches found in the neighborhoods, but only those whose positions relative to the considered match are similar are counted.
- Secondly, the test of similarity in relative positions is based on the relative distance (the value of r). Indeed, the similarity in relative positions is justified by the hypothesis that an affine transformation can approximate the change between the neighborhoods of the candidate match being considered. This assumption is reasonable only for a small neighborhood. Thus we should allow larger tolerance in distance differences for distant points, and this is exactly what our criterion does.
- Thirdly, the contribution of a candidate match $(\boldsymbol{n}_{1k}, \boldsymbol{n}_{2l})$ to the strength of the match $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$ is the exponential of the negative relative error r, which is strictly monotonically decreasing function of r. When r is very big, then $\exp(-r/\varepsilon_r) \to 0$, and the candidate match can be ignored. When $r \to 0$, i.e., the difference is very small, then $\exp(-r/\varepsilon_r) \to 1$, and the candidate will largely contribute to the match $(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j})$.
- Fourthly, if a point in the left image has several candidate matches in the right image, only the one which has smallest distance difference is accounted for, which is done by the "max" operator.
- Lastly, the contribution of each candidate match in the neighborhood is weighted by its distance to the match. The addition of "1" is only to prevent the over weight for very close points. In other words, a close candidate match gives more support to the match being considered than a distant one. This is also connected to the fact that an affine approximation is only reasonable for a small neighborhood.

The measure of matching support defined above, however, is not symmetric. That is, the strength of a match is possibly not the same if we reverse the role of the two images, i.e., possibly we have $S_M(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2j}) \neq S_M(\boldsymbol{m}_{2j}, \boldsymbol{m}_{1i})$. This occurs when several points $\boldsymbol{n}_{1k} \in \mathcal{N}(\boldsymbol{m}_{1i})$ are candidate matches of a single point $\boldsymbol{n}_{2l} \in \mathcal{N}(\boldsymbol{m}_{2j})$, as illustrated in Fig. 3 where $\boldsymbol{n}_{11}$ and $\boldsymbol{n}_{12}$ share the same point $\boldsymbol{n}_{21}$ as their candidate match. In our implementation, we have made the following modification in order to achieve the symmetry. Before computing the summation, if several points $\boldsymbol{n}_{1k} \in \mathcal{N}(\boldsymbol{m}_{1i})$ score the maximal value with the same point $\boldsymbol{n}_{2l} \in \mathcal{N}(\boldsymbol{m}_{2j})$, then only the point which gives the largest value is counted. This assures that the same pairing will be counted if we reverse the role of the two images.

Other heuristics can be integrated into the computation of the strength of a match. For example, if the angle of the rotation in the image plane is assumed to be less than Θ, then we can impose the following constraint: the angle between $\overrightarrow{m_{1i}n_{1k}}$ and $\overrightarrow{m_{2j}n_{2l}}$ must be less than Θ. In other words, for a candidate match (n_{1k}, n_{2l}) which does not satisfy the above constraint, its $\delta(m_{1i}, m_{2j}; n_{1k}, n_{2l})$ takes the value of zero.

In our implementation, R = one eighth of the image width, $c_{ij} = 1$, $\varepsilon_r = 0.3$ and $\Theta = 90°$.

5.2. Relaxation process

If we define the energy function as the sum of the strengths of all candidate matches, i.e.,

$$\mathcal{J} = \sum_{(m_{1i}, m_{2j})} S_M(m_{1i}, m_{2j}),$$

then the problem of disambiguating matches is equivalent to minimizing the energy function $\mathcal{J}$. The relaxation scheme is one approach to it. It is an iterative procedure, and can be formulated as follows:

iterate {
- compute the matching strength for each candidate match
- update the matches by minimizing the total energy

} **until the convergence of the energy**

After the correlation procedure, for each point in the first image, we have a set of candidate matches from the second image (the set is possibly nil); and in the same time we have also a set of candidate matches from the first image for each point in the second image. The last subsection has already explained how to compute the SM for each candidate match. As the definition of SM is now symmetric, we only need to compute SMs for the list of candidate matches in the left image and assign the values to the candidate matches in the right image, thus saving half of the computation.

There are several strategies for updating the matching in order to minimize the total energy. The first is the "*winner-take-all*", as exploited by Rosenfeld et al. [47], Zucker et al. [60], and Pollard et al. [44]. The method works as follows. At each iteration, any matches which have the highest matching strengths for both of the two image points that formed them are immediately chosen as "correct". That is, a match (m_{1i}, m_{2j}) is selected if its points (either m_{1i} or m_{2j}) have no higher matching-strength scores with any other matches they can form. Then, because of the uniqueness constraint, all other matches associated with the two points in each chosen match are eliminated from further consideration. This allows further matches, that were not previously either selected or eliminated, to be selected as correct provided they now have the highest matching strengths for both constituent points. This method proceeds as a steepest-descent approach, and is thus fast. However, it may get stuck easily at a bad local minimum.

The second is the "*loser-take-nothing*" [28]. The method works as follows. For each point in the first image, the candidate which has gained the weakest matching strength is eliminated. The process suppresses at most one candidate at each iteration until one and only one candidate is left for each point, finally achieving an unambiguous set of matches. Since the suppressed matches have gained the weakest support, they are very possibly not among the correct matches. This method thus proceeds as a slowest-descent approach, and is not efficient if a point has many candidate matches. Furthermore, this method is not symmetric for the two images: reversing the role of the two images may give different result.

We have developed a new update strategy, which we would like to call "*some-winners-take-all*". It differs from "winner-take-all", which is in fact *all-winners-take-all*, and works as follows. As with "winner-take-all", we consider all matches which have the highest matching strength for both of the two image points that formed them. We shall call such matches the *potential matches*, and denote them by $\{\mathcal{P}_i\}$. For $\{\mathcal{P}_i\}$, we construct two tables. The first, denoted by $\mathcal{T}_{\mathrm{SM}}$, saves the matching strength of each $\mathcal{P}_i$, and is then sorted in decreasing order. The second, denoted by $\mathcal{T}_{\mathrm{UA}}$, saves a value which indicates how unambiguous each $\mathcal{P}_i$ is. This is defined as

$$U_A = 1 - S_M^{(2)}/S_M^{(1)},$$

where $S_M^{(1)}$ is the SM of $\mathcal{P}_i$, and $S_M^{(2)}$ is the SM of the second best candidate match. Thus U_A is ranging from 1 (unambiguous) to 0 (ambiguous). The table $\mathcal{T}_{\mathrm{UA}}$ is also sorted in decreasing order. Finally, those potential matches $\mathcal{P}_i$ which are among both the first q percent of matches in $\mathcal{T}_{\mathrm{SM}}$ and the first q percent of matches in $\mathcal{T}_{\mathrm{UA}}$ are selected as correct matches. Thus, ambiguous potential matches will not be selected even they have high SM, and those having weak SM will not selected even they are unambiguous. We have therefore prevented the problem of evolve-too-soon-ness with "winner-take-all" while maintaining computational efficiency. If a candidate match does not receive any support ($S_M = 0$), it will be eliminated from further consideration. If $q = 100$, i.e., one hundred percent selection case, our method becomes "winner-take-all". Not that q must be larger than 50 in order to assure that at least one potential match will be selected at each iteration if there exist several potential matches. If $q < 50$, a premature stop may occur. In our implementation, q is set to 60.

Our algorithm necessarily converges, because if during one iteration there is no match selected, then the total energy will remain the same at the next iteration. The number of selected matches is evidently limited because the number of candidate matches is limited.

6. Robust estimation of the epipolar geometry

Using the set of matched points established in the previous step, one may then recover the so-called fundamental matrix. This is one of the most crucial steps. We will consider linear and nonlinear criteria and also exploit a robust technique to detect the outliers in the correspondences.

6.1. The linear criterion

Eq. (2) can be written as a linear and homogeneous equation in the 9 unknown coefficients of matrix $\boldsymbol{F}$:

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{f} = 0, \tag{7}$$

where

$$\boldsymbol{u} = [u_1u_2, v_1u_2, u_2, u_1v_2, v_1v_2, v_2, u_1, v_1, 1]^{\mathrm{T}},$$
$$\boldsymbol{f} = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^{\mathrm{T}}.$$

Thus we know that if we are given 8 matches we will be able, in general, to determine a unique solution for $\boldsymbol{F}$, defined up to a scale factor. This approach, known as the eight-point algorithm, was introduced by Longuet-Higgins [29] for solving the motion and structure from motion problem, and has been extensively studied in the literature [27, 30, 53, 55] for the computation of the *Essential matrix* E (see Section 3). It has proven to be very sensitive to noise.

In practice, we are given many more than 8 matches and we use a least-squares method to solve

$$\min_{\boldsymbol{F}} \sum_i (\tilde{\boldsymbol{m}}_{2i}^{\mathrm{T}}\boldsymbol{F}\tilde{\boldsymbol{m}}_{1i})^2, \tag{8}$$

which can be rewritten as:

$$\min_{\boldsymbol{f}} \|\boldsymbol{U}\boldsymbol{f}\|^2, \quad \text{where} \quad \boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{u}_n^{\mathrm{T}} \end{bmatrix}.$$

Several methods are possible to solve this problem. The first uses a closed-form solution via the linear equations by setting one of the coefficients of $\boldsymbol{F}$ to 1. The second solves the classical problem:

$$\min_{\boldsymbol{f}} \|\boldsymbol{U}\boldsymbol{f}\|^2 \quad \text{subject to } \|\boldsymbol{f}\| = 1. \tag{9}$$

The solution is the eigenvector of $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}$ associated with the smallest eigenvalue.

The advantage of the linear criterion is that it leads to a non-iterative computation method, however, we have found that it is quite sensitive to noise, even with a large set of data points. The two main reasons for this are:

- The constraint $\det(\boldsymbol{F}) = 0$ is not satisfied, which causes inconsistencies of the epipolar geometry near the epipoles.
- The criterion is not normalized, which causes a bias in the localization of the epipoles.

The reader is referred to [32] for a detailed study of the linear methods.

6.2. Minimizing the distances to epipolar lines

As it has been said, one of the drawbacks of the linear criterion method is that we do not take into account the fact that the rank of $\boldsymbol{F}$ is only two, and that $\boldsymbol{F}$ thus depends on only 7 parameters. This could be taken into account by parameterizing matrix $\boldsymbol{F}$ as follows:

$$\boldsymbol{F} = \begin{pmatrix} b & a & -ay - bx \\ -d & -c & cy + dx \\ dy' - bx' & cy' - ax' & -cyy' - dy'x + ayx' + bxx' \end{pmatrix}, \tag{10}$$

where the parameters (x, y) and (x', y') are the affine coordinate of the two epipoles, and the coefficients, of the homography between the two pencils of epipolar lines, which are the coefficients of the submatrix 2×2 obtained by suppressing the third line and the third column. However, nonlinear minimizations must be performed.

The first idea is to use a nonlinear criterion by minimizing:

$$\sum_i d^2(\tilde{\boldsymbol{m}}_{2i}, \boldsymbol{F}\tilde{\boldsymbol{m}}_{1i}),$$

where $d(\tilde{\boldsymbol{m}}_2, \boldsymbol{F}\tilde{\boldsymbol{m}}_1)$ is the Euclidean distance of point $\boldsymbol{m}_2$ to its epipolar line $\boldsymbol{F}\tilde{\boldsymbol{m}}_1$. It is given by

$$d(\tilde{\boldsymbol{m}}_2, \boldsymbol{F}\tilde{\boldsymbol{m}}_1) = \frac{|\tilde{\boldsymbol{m}}_2^{\mathrm{T}} \boldsymbol{F}\tilde{\boldsymbol{m}}_1|}{\sqrt{(\boldsymbol{F}\tilde{\boldsymbol{m}}_1)_1^2 + (\boldsymbol{F}\tilde{\boldsymbol{m}}_1)_2^2}},$$

where $(\boldsymbol{F}\tilde{\boldsymbol{m}}_1)_i$ is the ith component of vector $\boldsymbol{F}\tilde{\boldsymbol{m}}_1$. However, unlike the case of the linear criterion, the two images do not play a symmetric role. To obtain a consistent epipolar geometry, it is necessary and sufficient that by exchanging the two images, the fundamental matrix is changed to its transpose. This yields the following criterion:

$$\sum_i \left(d^2(\tilde{\boldsymbol{m}}_{2i}, \boldsymbol{F}\tilde{\boldsymbol{m}}_{1i}) + d^2(\tilde{\boldsymbol{m}}_{1i}, \boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{m}}_{2i})\right),$$

which operates simultaneously in the two images. Using the fact that $\tilde{\boldsymbol{m}}_2^{\mathrm{T}}\boldsymbol{F}\tilde{\boldsymbol{m}}_1 = \tilde{\boldsymbol{m}}_1^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{m}}_2$, it can be rewritten as:

$$\sum_i \left(\frac{1}{(\boldsymbol{F}\tilde{\boldsymbol{m}}_{1i})_1^2 + (\boldsymbol{F}\tilde{\boldsymbol{m}}_{1i})_2^2} + \frac{1}{(\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{m}}_{2i})_1^2 + (\boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{m}}_{2i})_2^2}\right)(\tilde{\boldsymbol{m}}_{2i}^{\mathrm{T}}\boldsymbol{F}\tilde{\boldsymbol{m}}_{1i})^2. \tag{11}$$

This criterion is also clearly normalized in the sense that it does not depend on the scale factor used to compute $\boldsymbol{F}$.

6.3. Taking into account possible outliers in the initial correspondences

In all matches established so far, as described in Section 5, we may find two types of *outliers* due to:

- *Bad locations.* In the estimation of the fundamental matrix, the location error of a point of interest is assumed to exhibit Gaussian behavior. This assumption is reasonable since the error in localization for most points of interest is small (within one or two pixels), but a few points are possibly incorrectly localized (more than three pixels). The latter points will severely degrade the accuracy of the estimation.
- *False matches.* In the establishment of correspondences, only heuristics have been used. Because the only geometric constraint, i.e., the epipolar constraint in terms of the *fundamental matrix*, is not yet available, many matches are possibly false. These will completely spoil the estimation process, and the final estimate of the fundamental matrix will be useless.

The outliers will severely affect the precision of the fundamental matrix if we directly apply the methods described above. In the following, we give a brief description of the two most popular robust methods: the *M-estimators* and the *Least Median of Squares* (LMedS) method.

Let r_i be the *residual* of the ith datum, i.e., the difference between the ith observation and its fitted value. The standard least-squares method tries to minimize $\sum_i r_i^2$, which is unstable if there are outliers present in the data. The M-estimators replace the squared residuals r_i^2 by another functions of the residuals, yielding

$$\min \sum_i \rho(r_i),$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero. For example, Huber [25] employed the squared error for small residuals and the absolute error for large residuals. The M-estimators can be implemented as a weighted least-squares problem. In [31,43], the following weight was used for the estimation of the epipolar geometry:

$$w_i = \begin{cases} 1, & |r_i| \leqslant \sigma, \\ \sigma/|r_i|, & \sigma < |r_i| \leqslant 3\sigma, \\ 0, & 3\sigma < |r_i|, \end{cases}$$

where σ is some estimated standard deviation of errors. This method was robust to outliers due to bad localization. It was, however, not robust to false matches.

The LMedS method estimates the parameters by solving the nonlinear minimization problem:

$$\min \underset{i}{\text{med}}\ r_i^2.$$

That is, the estimator must yield the smallest value for the median of squared residuals computed for the entire data set. It turns out that this method is very robust to false matches as well as outliers due to bad localization. Unlike the M-estimators, however, the LMedS problem cannot be reduced to a weighted least-squares problem. It is probably impossible to write down a straightforward formula for the LMedS estimator. It must be solved by a search in the space of possible estimates generated from the data. Since this space is too large, only a randomly chosen subset of data can be analyzed. The

algorithm which we have implemented for robustly estimating the fundamental matrix follows that structured in [48, Chapter 5], as outlined below.

Given n point correspondences: $\{(\boldsymbol{m}_{1i}, \boldsymbol{m}_{2i})\}$. A Monte Carlo type technique is used to draw m random subsamples of $p = 8$ different point correspondences. For each subsample, indexed by J, we determine the fundamental matrix $\boldsymbol{F}_J$. For each $\boldsymbol{F}_J$, we can determine the median of the squared residuals, denoted by M_J, with respect to the whole set of point correspondences, i.e.,

$$M_J = \underset{i=1,\ldots,n}{\text{med}} [d^2(\tilde{\boldsymbol{m}}_{2i}, \boldsymbol{F}_J \tilde{\boldsymbol{m}}_{1i}) + d^2(\tilde{\boldsymbol{m}}_{1i}, \boldsymbol{F}_J^{\mathrm{T}} \tilde{\boldsymbol{m}}_{2i})].$$

We retain the estimate $\boldsymbol{F}_J$ for which M_J is minimal among all m M_J's. The question now is: *How do we determine m*? A subsample is "good" if it consists of p good correspondences. Assuming that the whole set of correspondences may contain up to a fraction ε of outliers, the probability that at least one of the m subsamples is good is given by

$$P = 1 - [1 - (1 - \varepsilon)^p]^m. \tag{12}$$

By requiring that P must be near 1, one can determine m for given values of p and ε. In our implementation, we assume $\varepsilon = 40\%$ and require $P = 0.99$, thus $m = 272$. Note that the algorithm can be speeded up considerably by means of parallel computing, because the processing for each subsample can be done independently.

As noted in [48], the LMedS *efficiency* is poor in the presence of Gaussian noise. The efficiency of a method is defined as the ratio between the lowest achievable variance for the estimated parameters and the actual variance provided by the given method. To compensate for this deficiency, we further carry out a weighted least-squares procedure. The *robust standard deviation* estimate is given by

$$\hat{\sigma} = 1.4826[1 + 5/(n - p)]\sqrt{M_J},$$

where M_J is the minimal median. The reader is referred to [48, page 202] for the explanation of these magic numbers. Based on $\hat{\sigma}$, we can assign a weight for each correspondence:

$$w_i = \begin{cases} 1, & \text{if } r_i^2 \leqslant (2.5\hat{\sigma})^2, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$r_i^2 = d^2(\tilde{\boldsymbol{m}}_{2i}, \boldsymbol{F}\tilde{\boldsymbol{m}}_{1i}) + d^2(\tilde{\boldsymbol{m}}_{1i}, \boldsymbol{F}^{\mathrm{T}}\tilde{\boldsymbol{m}}_{2i}).$$

The correspondences having $w_i = 0$ are outliers and should not be further taken into account. The fundamental matrix $\boldsymbol{F}$ is finally estimated by solving the weighted least-squares problem:

$$\min \sum_i w_i r_i^2.$$

We have thus robustly estimated the fundamental matrix because outliers have been detected and discarded by the LMedS method.

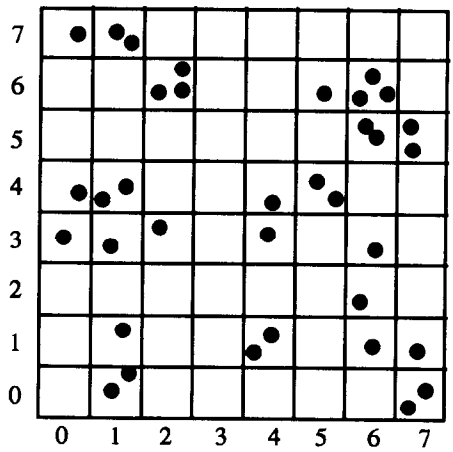

Fig. 4. Illustration of a bucketing technique.

As said previously, computational efficiency of the LMedS method can be achieved by applying a Monte Carlo type technique. However, the eight points of a subsample thus generated may be very close to each other. Such a situation should be avoided because the estimation of the epipolar geometry from such points is highly instable and the result is useless. It is a waste of time to evaluate such a subsample. In order to achieve higher stability and efficiency, we develop a *regularly random selection method* based on bucketing techniques, which works as follows. We first calculate the `min` and `max` of the coordinates of the points in the first image. The region is then evenly divided into $b \times b$ buckets (see Fig. 4). To each bucket is attached a set of points, and indirectly a set of matches, which fall in it. The buckets having no matches attached are excluded. To generate a subsample of 8 points, we first randomly select 8 mutually different buckets, and then randomly choose one match in each selected bucket.

One question remains: How many subsamples are required? If we assume that bad matches are uniformly distributed in space, and if each bucket has the same number of matches and the random selection is uniform, the formula (12) still holds. However,

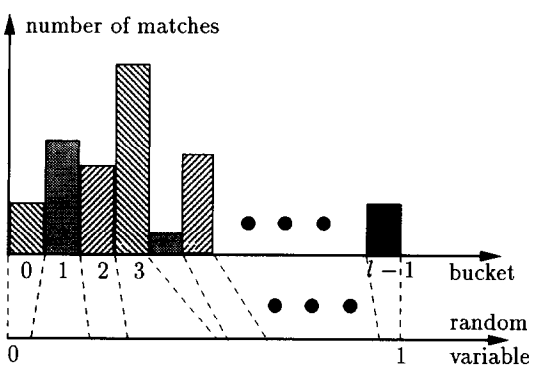

Fig. 5. Interval and bucket mapping.

the number of matches in one bucket may be quite different from that in another. As a result, a match belonging to a bucket having fewer matches has a higher probability to be selected. It is thus preferred that a bucket having many matches has a higher probability to be selected than a bucket having few matches, in order that each match has almost the same probability to be selected. This can be realized by the following procedure. If we have in total l buckets, we divide $[0, 1]$ into l intervals such that the width of the ith interval is equal to $n_i / \sum_i n_i$, where n_i is the number of matches attached to the ith bucket (see Fig. 5). During the bucket selection procedure, a number, produced by a $[0, 1]$ uniform random generator, falling in the ith interval implies that the ith bucket is selected.

In our implementation, $b = 8$.

7. Stereo matching

Once the fundamental matrix has been determined robustly, we use it to establish a new set of correspondences using a correlation based approach that takes into account the recovered epipolar geometry (i.e., epipolar constraint).

The matching approach that has been developed is a slightly modified version of the initial matching process (Section 4). For a feature point in the first image, and in order to find possible matching partners not too far from the epipolar line in the second image, we place a narrow band of width 2ε pixels centered on this epipolar line and find the points that lie within the band. The value of ε is chosen to be $3.8\bar{d}$ for a probability of 95%, where $\bar{d}$ is the root of mean squares of distances between the points and their epipolar lines defined by the recovered fundamental matrix, i.e., $\bar{d} = \sqrt{\sum_i w_i r_i^2 / \sum_i w_i}$. The same constraints as in Section 4 are then applied to select the most consistent matches, except that the constraint on the disparity (defined by the search window) is replaced by the epipolar constraint just described.

If needed, we can refine the estimation of the fundamental matrix using all correspondences established at this point. The number of correspondences found in this step is usually larger.

8. Experimental results

The proposed algorithm has been tested on two dozen image pairs, and good results have been obtained. Different types of scenes have been used, such as indoor, rocks, road, and textured dummy scenes. Due to space limitation, we provide in this paper the matching results of six image pairs, which are labeled **bust**, **road**, **valley**, **rotation**, **trunk** and **tracing**. The last image pair has been generated by a ray-tracing algorithm, and contains many repetitive patterns. All others are real images. The image resolution is 512×512, except for scene **road**, whose resolution is 512×470. All parameters used in the algorithm are the same for all image pairs with one exception, and are as were specified in the previous sections. The exception was done for image pair **rotation**. Because of its large disparity in y-direction, the size of correlation search

Table 1
Summary of the matching results with the correlation, relaxation and robust methods: Numbers of total and bad matches

Scene name	No. of pts. (left–right)	Correlation (total/bad)	Relaxation (total/bad)	LMedS (detected)	Stereo (matches)	No. of iterations
bust	512–512	103/4	97/0	11	93	11
road	367–389	53/12	52/7	5	48	12
valley	395–512	235/14	248/4	15	241	16
rotation	288–255	89/31	102/11	15	88	12
trunk	377–355	117/25	127/20	22	118	15
tracing	383–352	143/*lots*	248/19	31	226	20

window, which is set to 257×257 by default, is not big enough, and we have set it to 257×301.

In order to show the performance of the relaxation procedure, we also provide the matching results given by the correlation technique. The correlation results are obtained as follows. We perform the correlation twice by reversing the roles of the two images (i.e., from left to right, and then from right to left) and consider as valid only those matches for which the reverse correlation has fallen on the initial point in the first image. More precisely, for a given point $\boldsymbol{m}_1$ in the left image, let the match candidate with the highest correlation score be $\boldsymbol{m}_2$ through a left-to-right correlation. Before validating the match, we perform a right-to-left correlation. If the match candidate with the highest correlation score for $\boldsymbol{m}_2$ is again $\boldsymbol{m}_1$, then this match will be validated; otherwise, it will be rejected. The two images thus play a symmetric role. This validity test allows us to reduce greatly the probability of error.

The results are shown in Figs. 6 through 23, and are summarized in Table 1. For each scene, three figures are provided, showing the matching results by correlation, relaxation and stereo. By stereo is meant that the epipolar geometry estimated by the robust method LMedS is used in matching, as described in Section 7. In each figure, two pictures (the left and right images) are shown. Points are indicated by a cross, and matched points are given the same number. The epipolar lines are drawn on the images to illustrate the difference between the epipolar geometry estimated from all matches found by relaxation and that estimated by the LMedS. In Table 1, both the total number of matches and the number of bad matches found by correlation and relaxation are provided, together with the number of outliers detected by the LMedS and the number of matches found by stereo. Here, we only count false matches as bad matches, and do not take into account those matches due to bad location. The LMedS actually detects all those bad matches except for scene **road**, and the matches due to bad locations. Sometimes, the LMedS also rejects a few good matches whose error measures go slightly beyond the threshold defined in Section 6.3. The last column shows the number of iterations conducted during the relaxation process. The word *lots* in the last row means that the number of bad matches is very high (more than 50% and not checked carefully) when the correlation technique was used for scene **tracing**. Just for information, the first column gives the number of points of interest for each of the

Table 2
Comparison of the normalized residuals before and after discarding outliers by LMedS (in pixels)

	bust	road	valley	rotation	trunk	tracing
before	0.6	5.1	3.8	9.2	13.0	8.5
after	0.3	0.8	0.7	0.5	1.2	0.4

first and second images.

In Table 2, the normalized residuals before and after discarding outliers by LMedS are given, which show the improvement achieved by incorporating an outlier detection module. A normalized residual is approximately equal to the average distance between a point and its epipolar line.

For scene **bust** (Figs. 6–8), the images were taken by two cameras with parallel optical axes placed almost horizontally. There were a few false matches when using the correlation technique, e.g., match 76 (on the right part of the images in Fig. 6). The matches found with the relaxation are all correct. The epipolar geometry computed from these matches is shown by the epipolar lines in Fig. 7. The average distance between a point and its epipolar line is 0.6 pixels. Consider that the precision of a point extracted by the corner detector described in Section 4.1 is in the order of one pixel, the epipolar geometry estimated is quite good. Two good matches found by correlation are missed by relaxation: matches 4 and 15 in the top part of images in Fig. 6. This is because these matches are isolated and too far away from other matches to gain any support. Recall that the radius of the neighborhood disc R is set to the eighth of the image width. If we increase the value of R, we can recover these two matches. Although all matches are correct, LMedS still detects 11 matches as outliers, and the average distance between a point and its epipolar line is now reduced to 0.3 pixels. Comparing the epipolar geometry illustrated in Fig. 8 with that in Fig. 7, one can hardly find any difference.

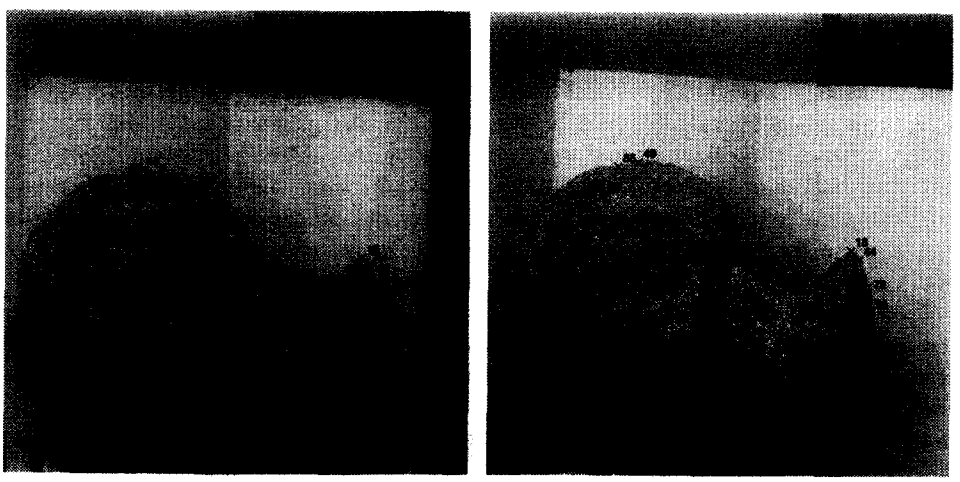

Fig. 6. Scene **bust**: Matching result with the correlation technique.

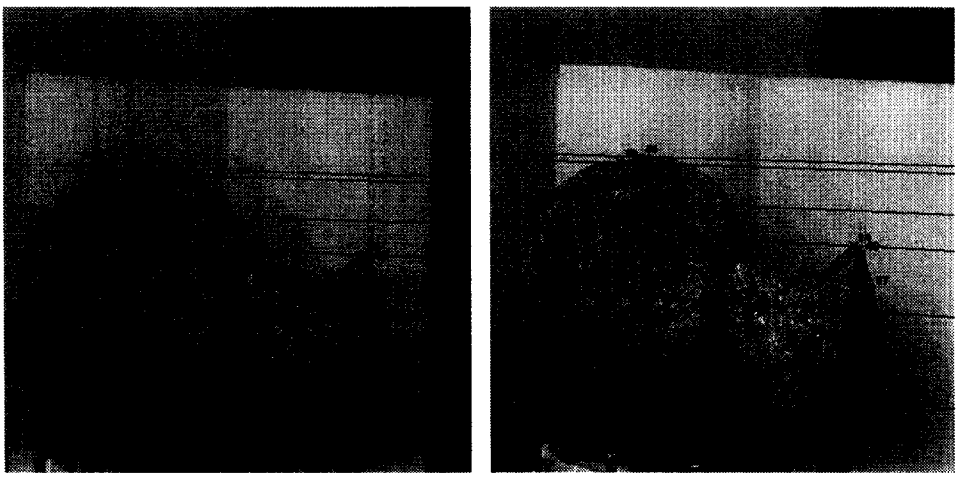

Fig. 7. Scene **bust**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

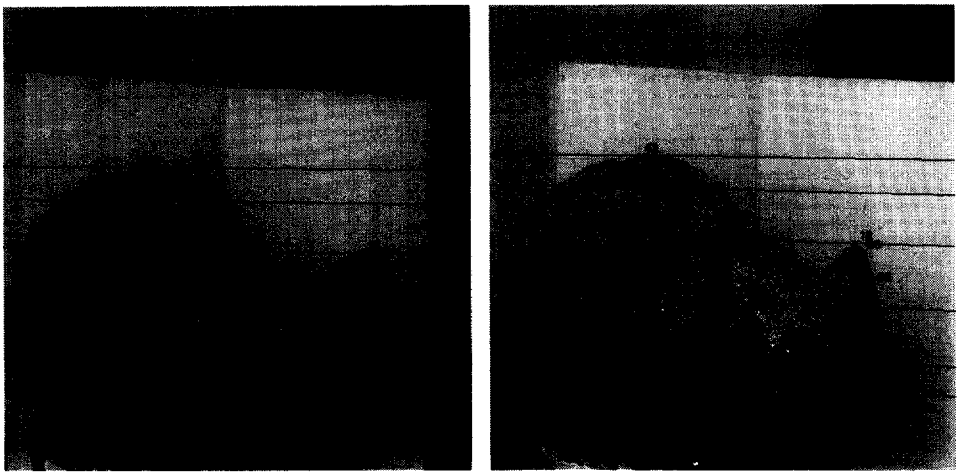

Fig. 8. Scene **bust**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

For scene **road** (Figs. 9–11), the two images were taken by a single camera mounted on a moving vehicle (this is thus a motion sequence). The vehicle moves forward on the right lane, and the epipolar lines are thus expected to intersect to each other at a point near the center of the image. Fig. 9 shows the matched points recovered by using the correlation technique. One can notice that some points have not been correctly matched: matches 5, 21 and 23, to name a few. After carrying out the relaxation procedure, several false matches have been corrected, but a few still persist, e.g., match 22 in Fig. 10. The epipolar geometry estimated, as shown by the epipolar lines in Fig. 10, is not

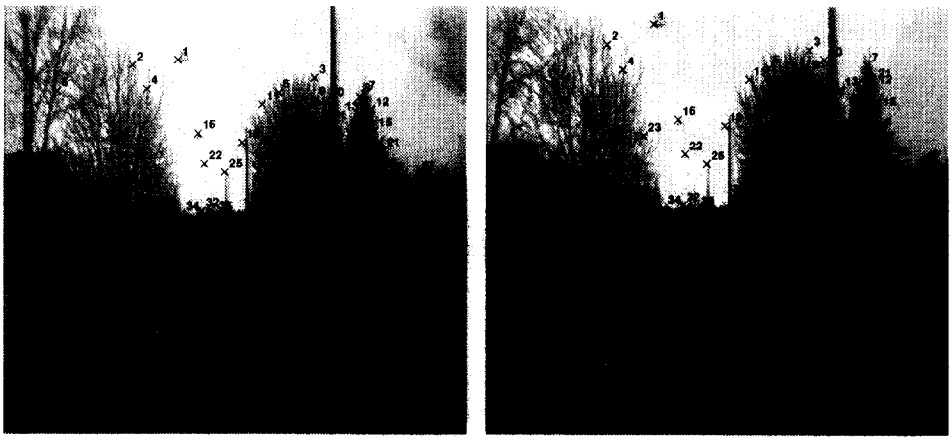

Fig. 9. Scene **road**: Matching result with the correlation technique.

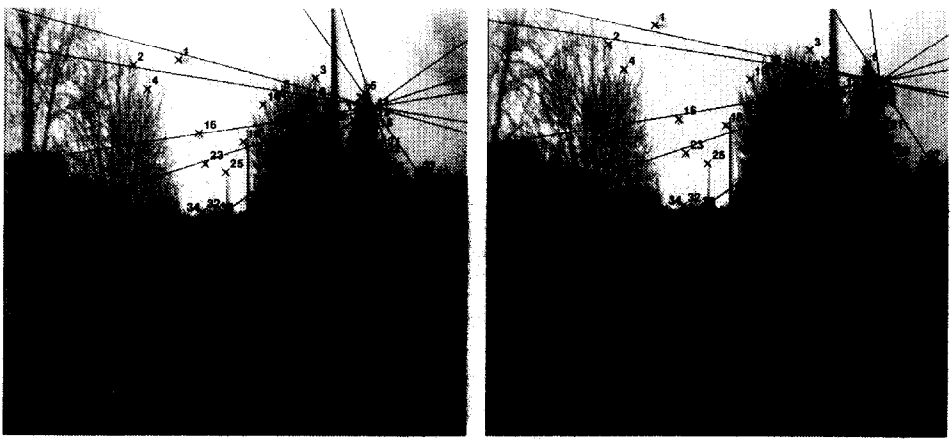

Fig. 10. Scene **road**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

correct, and the average distance between a point and its epipolar line is 5.1 pixels, which is quite large. The LMedS has detected and rejected 5 outliers (3 false matches and 2 not very well located points). This significantly changes the epipolar geometry, in particular, the positions of the epipoles, as shown in Fig. 11. The average distance is now 0.8 pixels. The attentive reader may have noticed that the number of outliers detected, 5, is less than the number of false matches, 7 (see Table 1). In fact, there are four matches (49, 50, 51 and 52 in Fig. 10) on the lane marker which are not correct. However, they have not been detected by the LMedS because they almost lie on the same epipolar line (see Fig. 11). As the criterion used in outlier detection is the epipolar constraint, false matches lying on the epipolar line cannot be detected by our LMedS method.

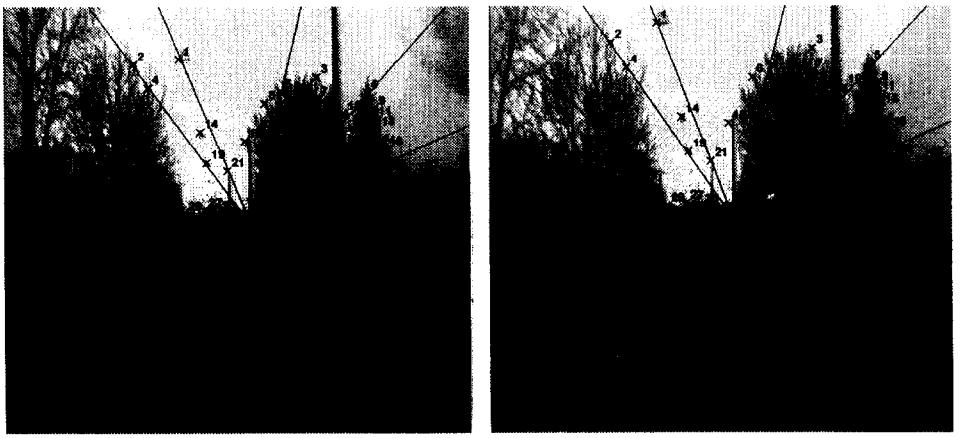

Fig. 11. Scene **road**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

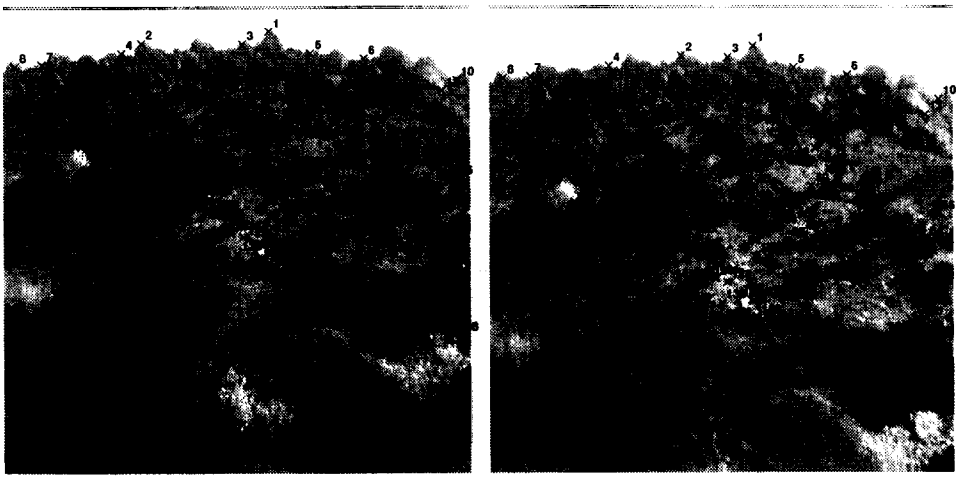

Fig. 12. Scene **valley**: Matching result with the correlation technique.

For scene **valley** (Figs. 12–14), the two images were taken by two cameras placed one above the other. The epipolar lines are thus expected to be almost parallel and oriented vertically. The scene is composed of rocks with different sizes. As the scene is rather textured, the correlation technique works reasonably well: 235 matches have been found, 14 of which are false (see Fig. 12). After the relaxation process, even more matches (248) have been recovered, and only four of them are false. The false matches are those labeled 184, 54, 49 and 214 in Fig. 13. Although there are only four false matches, the epipolar geometry estimated is completely wrong, as shown by the epipolar lines in Fig. 13, and the average distance between a point and its epipolar line is 3.8 pixels, which is large. The LMedS has detected all these false matches plus several not

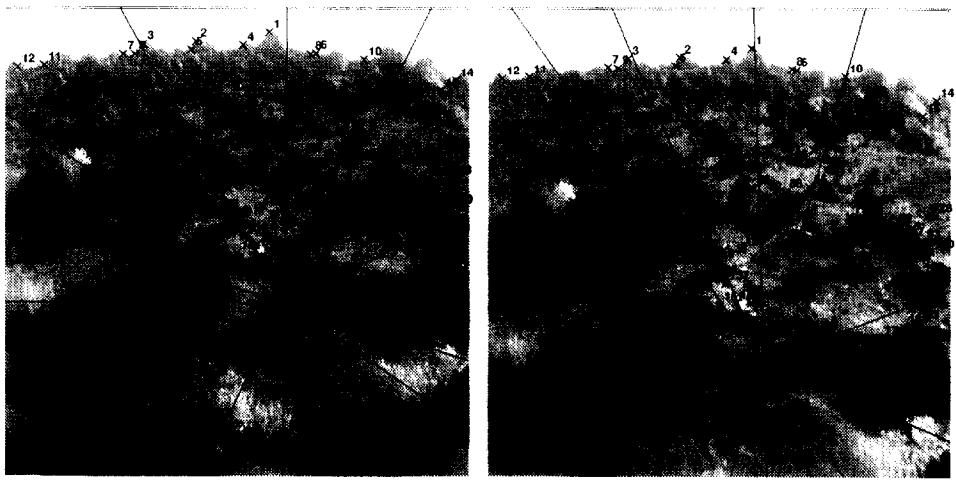

Fig. 13. Scene **valley**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

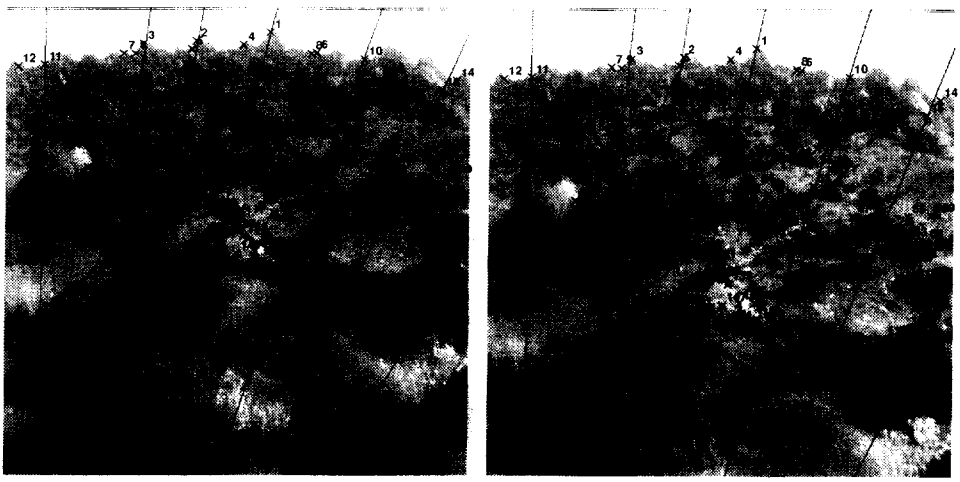

Fig. 14. Scene **valley**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

very well located ones. This completely changes the epipolar geometry, as shown in Fig. 14. The average distance is now 0.7 pixels.

For scene **rotation** (Figs. 15–17), the two images were taken by the same camera, but there is a rotation around the optical center and a tilt translation between the two positions. In spite of the image distortion due to the rotation, the correlation works reasonably except for the points on the grid against the wall, where the repetitive patterns make matching by correlation extremely difficult (see Fig. 15): among 89 matches recovered, 31 are false. Because of the use of contextual (neighboring) information, the

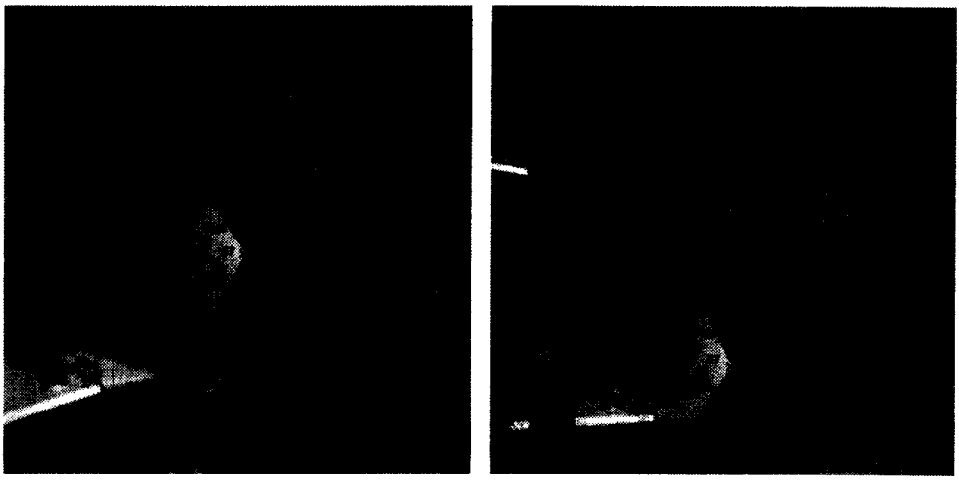

Fig. 15. Scene **rotation**: Matching result with the correlation technique.

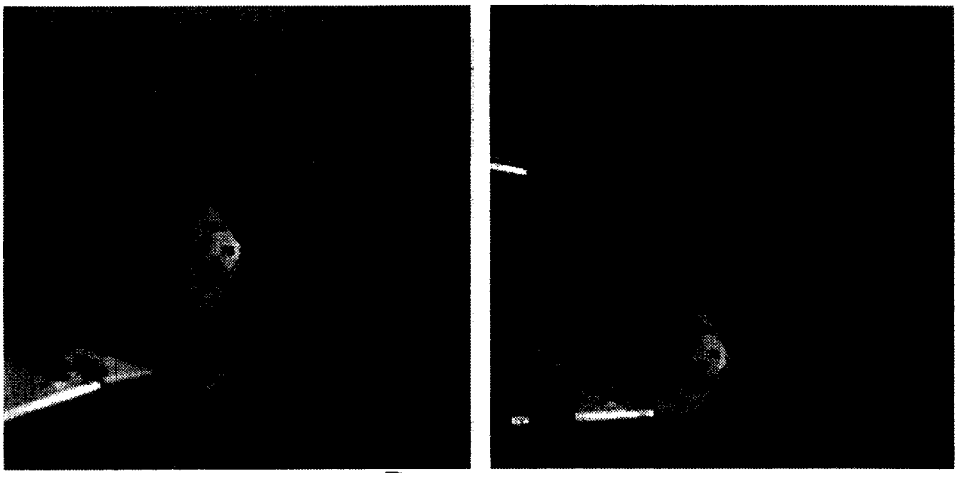

Fig. 16. Scene **rotation**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

relaxation technique has produced a much better matching result (see Fig. 16). Among 102 matches recovered, only 11 are false, and almost all matches on the grid are correct. The LMedS has detected all the false matches, and the epipolar geometry estimated is illustrated by lines in Fig. 17. The average distance between a point and its epipolar line is reduced from 9.2 pixels to 0.5.

For scene **trunk** (Figs. 18–20), the two images were taken by two cameras placed side by side. The two cameras have quite different focal lengths, as can be noticed by the change of the trunks' size. Again, despite the scale difference, our method works well. One can also notice the drastic change in the epipolar geometry estimated before

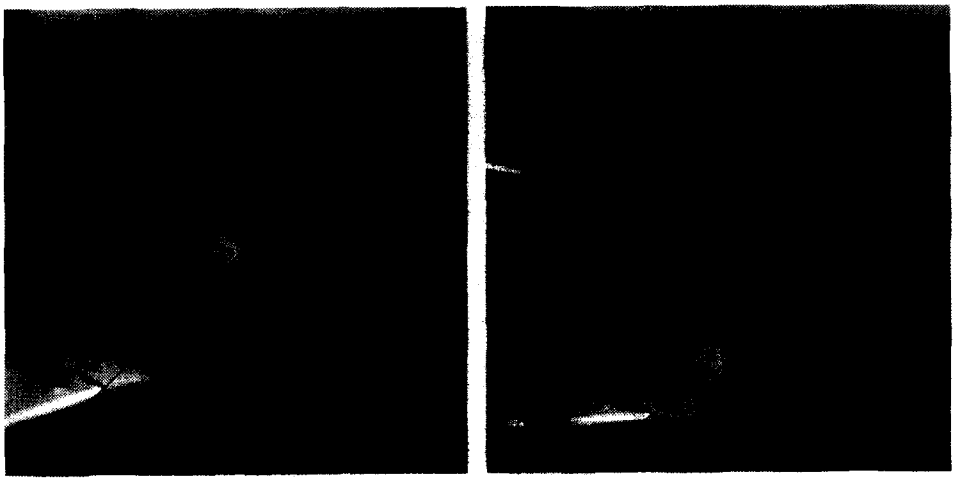

Fig. 17. Scene **rotation**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

Fig. 18. Scene **trunk**: Matching result with the correlation technique.

(Fig. 19) and after (Fig. 20) applying the LMedS technique for outlier detection. The average distance between a point and its epipolar line is reduced from 13 pixels to 1.2.

Finally, the two images in scene **tracing** (Figs. 21–23) were generated by a ray-tracing technique simulating two cameras placed diagonally. The numerous repetitive patterns make the matching by correlation almost impossible. Indeed, among 143 matches found, more than a half are false (Fig. 21). This implies that if we apply the LMedS at this stage we cannot obtain any useful result. Using the relaxation technique, we have obtained a very good matching result: 248 matches have been recovered, and only 19 are false (Fig. 22). These false matches have been all detected by the LMedS, and

Fig. 19. Scene **trunk**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

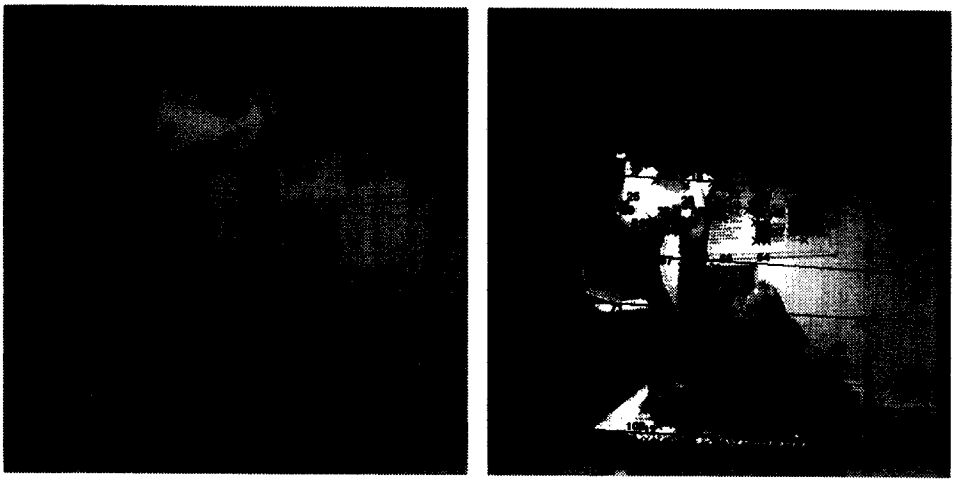

Fig. 20. Scene **trunk**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

the epipolar geometry has been correctly estimated, as shown in Fig. 23. The average distance between a point and its epipolar line is reduced from 8.5 pixels in Fig. 22 to 0.4 pixels in Fig. 23. The fundamental matrix used for generating the two images is:

$$\begin{bmatrix} 6.440951e{-}07 & 5.203664e{-}06 & 1.658593e{-}02 \\ -4.065228e{-}06 & 7.716572e{-}07 & 1.798488e{-}02 \\ -1.821295e{-}02 & -1.834903e{-}02 & 1 \end{bmatrix}$$

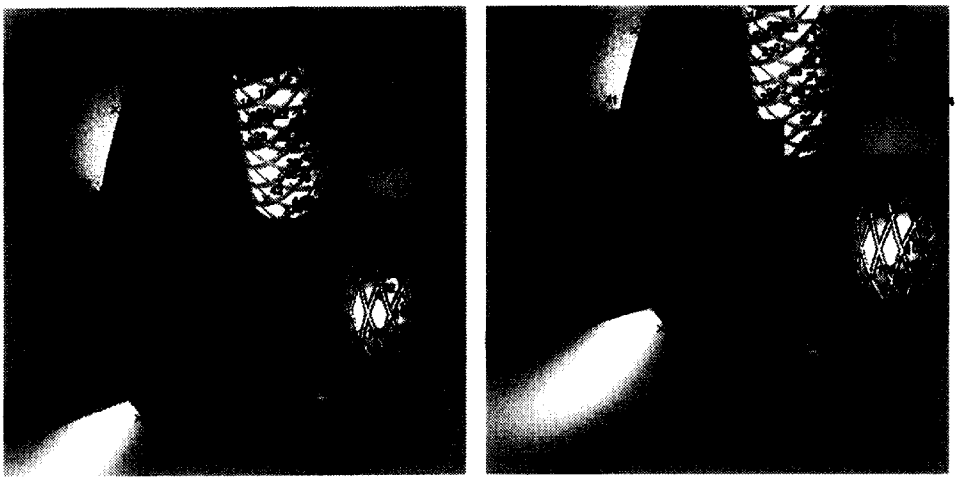

Fig. 21. Scene **tracing**: Matching result with the correlation technique.

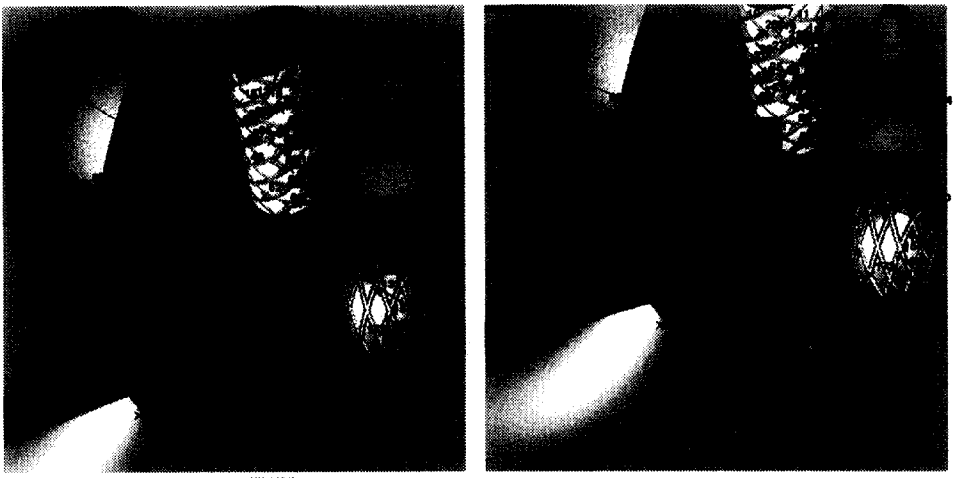

Fig. 22. Scene **tracing**: Matching result with the relaxation technique and the epipolar geometry recovered using all matched points.

and the estimated one is:

$$\begin{bmatrix} 6.455367e{-}07 & 5.146858e{-}06 & 1.622137e{-}02 \\ -4.012881e{-}06 & 7.702527e\ 07 & 1.775179e{-}02 \\ -1.785370e{-}02 & -1.811788e{-}02 & 1 \end{bmatrix}$$

For comparison, we have normalized the last element to 1. The difference between them is very small.

In Table 3, we summarize the CPU time in seconds spent in each step on a Spar 10

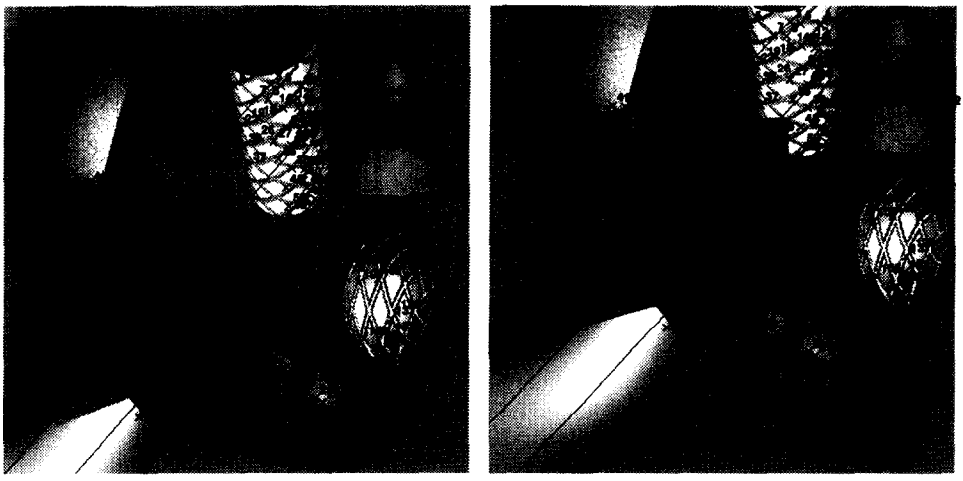

Fig. 23. Scene **tracing**: The epipolar geometry recovered after discarding outliers and the matched points found using the epipolar constraint.

Table 3
Computation time in each step: CPU time in seconds on a Sparc 10 Workstation

Scene	Correlation	Relaxation	LMedS	Stereo
bust	16.9	5.9	2.2	1.8
road	8.4	1.9	1.5	2.0
valley	8.6	9.5	3.3	1.6
rotation	2.8	2.1	2.0	0.6
trunk	6.3	4.9	2.1	2.0
tracing	6.1	76	3.2	1.1

Workstation. The time shown is merely indicative, as we have not tried to optimize our codes. The time for correlation depends on the scene complexity, and essentially the number of points. The time for relaxation depends on the degree of matching ambiguity. If there are not many matching candidates, like in scene **road**, the time is very short. In scene **tracing**, however, the time spent in relaxation is extremely high (76 seconds!) because each point may have as many as 40 matching candidates. The time for LMedS is almost constant, around 2 seconds. The time for stereo is around 1.6 seconds. Thus, in general, the most time consuming step is the correlation.

9. Conclusion

We have proposed in this paper a robust approach to image matching by exploiting the only geometric constraint, namely, the epipolar constraint. The presence of the epipolar geometry between two images is now well known, and can be described by the fundamental matrix. The idea is to use some heuristic techniques (correlation methods

in our particular implementation) to find an initial set of matches, and then use a very robust technique—the Least Median of Squares (LMedS)—to discard outliers in this set. The LMedS aims at finding the fundamental matrix, by searching in the parameter space, which minimizes the median of the squared errors. An error is quantified by the distance of a point to its corresponding epipolar line. This method has been tested with a large number of real images (indoor scenes, outdoor scenes, dummy mannequins, etc.). From the experiments we have carried out, our approach allows about 40% of false matches in the initial set of matches. The fundamental matrix can finally be accurately estimated from the good points.

There still exist a number of ways to improve our algorithm:

- *Accuracy.* The precision of the final estimation of the fundamental matrix depends tightly on those of the matched points. To have a better estimation of the fundamental matrix, we should increase the accuracy of matched points. One possibility is to use subpixel-precision corner detector whenever possible. For example, if we are working in an indoor environment, we may use corner detectors such as [8, 11, 46]. Another possibility is to apply subpixel-precision correlation techniques. The idea is to compute the correlation score for each point in the neighborhood of the match, and fit a parametric surface, e.g., a paraboloid, to the correlation scores, and eventually compute the position with the largest correlation score.
- *Stability.* The stability of our algorithm is directly related to that of finding the fundamental matrix, which is studied in [33]. One of the most stringent situations is that all points are located close to a critical surface, for example, a plane. If the points are almost on a plane, it is better to describe their relation between two views by a homography instead of a fundamental matrix. This will not change considerably the structure of our algorithm. The question is when to switch from using a fundamental matrix to using a homography, and vice versa.
- *Ambiguity.* Using the epipolar constraint alone does not allow to find unambiguous matches between two views. If a false match is occasionally aligned with the epipolar line, it will not be detected by our algorithm because only the epipolar constraint is used for outlier rejection. This of course does not disturb the estimation of the fundamental matrix. A good stereo matching algorithm, such as [34, 44], should be used to exploit other constraints like a disparity gradient limit.

Acknowledgment

The authors thank Thierry Blaszka for the corner detection code, Cyril Zeller for the correlation code and various discussions, and Charles Rothwell for checking the English.

References

[1] J. Aggarwal and N. Nandhakumar, On the computation of motion from sequences of images—a review, *Proc. IEEE* **76** (8) (1988) 917–935.

[2] H. Asada and M. Brady, The curvature primal sketch, *IEEE Trans. Pattern Anal. Mach. Intell.* **8** (1986) 2–14.

[3] S. Ayer, P. Schroeter and J. Bigün, Segmentation of moving objects by robust motion parameter estimation over multiple frames, in: J.-O. Eklundh, ed., *Proceedings Third European Conference on Computer Vision* **II**, Stockholm, Sweden (1994) 316–327.

[4] D.H. Ballard and C.M. Brown, *Computer Vision* (Prentice-Hall, Englewood Cliffs, NJ, 1982).

[5] S. Barnard and W. Thompson, Disparity analysis of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **2** (4) (1980) 333–340.

[6] J. Cheng and T. Huang, Image registration by matching relational structures, *Pattern Recogn.* **17** (1) (1984) 149–159.

[7] C.-H. Chou and Y.-C. Chen, Moment-preserving pattern matching, *Pattern Recogn.* **23** (5) (1990) 461–474.

[8] R. Deriche and T. Blaszka, Recovering and characterizing image features using an efficient model based approach, in: *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, New York (1993).

[9] R. Deriche and O. Faugeras, 2D-curves matching using high curvatures points: applications to stereovision, in: *Proceedings 10th International Conference on Pattern Recognition* **1**, Atlantic City, NJ (1990) 240–242.

[10] R. Deriche and O. Faugeras, Tracking line segments, in: O. Faugeras, ed., *Proceedings First European Conference on Computer Vision*, Antibes, France (1990) 259–268.

[11] R. Deriche and G. Giraudon, A computational approach for corner and vertex detection, *Int. J. Comput. Vision* **10** (2) (1993) 101–124.

[12] R. Deriche, Z. Zhang, Q.-T. Luong and O. Faugeras, Robust recovery of the epipolar geometry for an uncalibrated stereo rig, in: *Proceedings Third European Conference on Computer Vision* **I**, Stockholm, Sweden (1994) 567–576.

[13] O. Faugeras, F. Lustman and G. Toscani, Motion and structure from motion from point and line matches, in: *Proceedings First International Conference on Computer Vision*, London (1987) 25–34.

[14] P. Fua, A parallel stereo algorithm that produces dense depth maps and preserves image features, *Mach. Vision Appl.* **6** (1) (1993).

[15] A. Goshtasby, S.H. Gage and J.F. Bartholic, A two-stage cross correlation approach to template matching, *IEEE Trans. Pattern Anal. Mach. Intell.* **6** (3) (1984) 374–378.

[16] M. Hannah, A system for digital stereo image matching, *Photogrammetric Eng. Remote Sensing* **55** (12) (1989) 1765–1770.

[17] R. Haralick, Computer vision theory: the lack thereof, *Comput. Vision Graph. Image Process.* **36** (1986) 372–386.

[18] R. Haralick et al., Pose estimation from corresponding point data, *IEEE Trans. Syst. Man Cybernet.* **19** (6) (1989) 1426–1446.

[19] C. Harris, Determination of ego-motion from matched points, in: *Proceedings Alvey Conference* (1987).

[20] C. Harris and M. Stephens, A combined corner and edge detector, in: *Proceedings Alvey Conference* (1988) 189–192.

[21] R. Horaud and F. Veillon, Finding geometric and relational structures in an image, in: *Proceedings First European Conference on Computer Vision*, Antibes, France (1990) 374–384.

[22] R. Horaud and T. Skordas, Stereo correspondence through feature grouping and maximal cliques, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (11) (1989) 1168–1180.

[23] T. Huang and O. Faugeras, Some properties of the E matrix in two-view motion estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (12) (1989) 1310–1312.

[24] T. Huang and A. Netravali, Motion and structure from feature correspondences: a review, *Proc. IEEE* **82** (2) (1994) 252–268.

[25] P. Huber, *Robust Statistics* (John Wiley, New York, 1981).

[26] R. Kumar and A. Hanson, Analysis of different robust methods for pose refinement, in: *Proceedings International Workshop on Robust Computer Vision*, Seattle, WA (1990) 167–182.

[27] C.-H. Lee, Time-varying images: the effect of finite resolution on uniqueness, *Comput. Vision Graph. Image Process. Image Understanding* **54** (3) (1991) 325–332.

[28] S. Li, Inexact matching of 3D surfaces, Technical Report VSSP-TR-3/90, Vision Speech & Signal Processing, Department of Electronic and Electrical Engineering, University of Surrey, Guildford, UK (1990).

[29] H. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **293** (1981) 133–135.

[30] H. Longuet-Higgins, The reconstruction of a scene from two projections—configurations that defeat the 8-point algorithm, in: *Proceedings First Conference on Artificial Intelligence Applications*, Denver, CO (1984) 395–397.

[31] Q.-T. Luong, *Matrice fondamentale et calibration visuelle sur l'environnement: vers une plus grande autonomie des systèmes robotiques*, Dissertation, University of Paris XI, Orsay, Paris (1992).

[32] Q.-T. Luong, R. Deriche, O. Faugeras and T. Papadopoulo, On determining the fundamental matrix: analysis of different methods and experimental results, Rapport de Recherche 1894, INRIA Sophia-Antipolis, France (1993).

[33] Q.-T. Luong and O. Faugeras, A stability analysis of the fundamental matrix, in: J.-O. Eklundh, ed., *Proceedings Third European Conference on Computer Vision*, Stockholm, Sweden (1994) 577–588.

[34] R. Ma, M. Thonnat and M. Berthod, An adjustment-free stereo matching algorithm, in: J. Illingworth, ed., *Proceedings British Machine Vision Conference*, University of Surrey, Guildford, UK (1993) 609–618.

[35] H. Maître and Y. Wu, Improving dynamic programming to solve image registration, *Pattern Recogn.* **20** (4) (1987) 443–462.

[36] S. Maybank and O. Faugeras, A theory of self-calibration of a moving camera, *Int. J. Comput. Vision* **8** (2) (1992) 123–152.

[37] G. Medioni and Y. Yasumuto, Corner detection and curve representation using cubic b-spline, in: *Proceedings International Conference on Robotics and Automation*, San Francisco, CA (1986) 764–769.

[38] P. Meer, D. Mintz, A. Rosenfeld and D. Kim, Robust regression methods for computer vision: a review, *Int. J. Comput. Vision* **6** (1) (1991) 59–70.

[39] H. Nagel, Image sequences—ten (octal) years—from phenomenology towards a theoretical foundation, in: *Proceedings 8th International Conference on Pattern Recognition*, Paris, France (1986) 1174–1185.

[40] E. Nishimura, G. Xu and S. Tsuji, Motion segmentation and correspondence using epipolar constraint, in: *Proceedings 1st Asian Conference on Computer Vision*, Osaka, Japan (1993) 199–204.

[41] J. Noble, Finding corners, *Image Vision Comput.* **6** (1988) 121–128.

[42] J.-M. Odobez and P. Bouthemy, Robust multiresolution estimation of parametric motion models applied to complex scenes, Publication Interne 788, IRISA-INRIA Rennes, France (1994).

[43] S. Olsen, Epipolar line estimation, in: *Proceedings Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy (1992) 307–311.

[44] S. Pollard, J. Mayhew and J. Frisby, PMF: a stereo correspondence algorithm using a disparity gradient limit, *Perception* **14** (1985) 449–470.

[45] B. Radig, Image sequence analysis using relational structures, *Pattern Recogn.* **17** (1) (1984) 161–167.

[46] K. Rohr, Modelling and identification of characteristic intensity variations, *Image Vision Comput.* **10** (2) (1992) 66–76.

[47] A. Rosenfeld, R. Hummel and S. Zucker, Scene labeling by relaxation operations, *IEEE Trans. Syst. Man Cybernet.* **6** (4) (1976) 420–433.

[48] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection* (John Wiley, New York, 1987).

[49] L. Shapiro and M. Brady, Rejecting outliers and estimating errors in an orthogonal regression framework, Tech. Report OUEL 1974/93, Department of Engineering Science, University of Oxford (1993).

[50] L. Shapiro and R. Haralick, Structural description and inexact matching, *IEEE Trans. Pattern Anal. Mach. Intell.* **3** (1981) 504–519.

[51] W. Thompson, P. Lechleider and E. Stuck, Detecting moving objects using the rigidity constraint, *IEEE Trans. Pattern Anal. Mach. Intell.* **15** (2) (1993) 162–166.

[52] P. Torr and D. Murray, Outlier detection and motion segmentation, in: *Sensor Fusion* **VI**, SPIE Vol. 2059, Boston, MA (1993) 432–443.

[53] R. Tsai and T. Huang, Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surface, *IEEE Trans. Pattern Anal. Mach. Intell.* **6** (1) (1984) 13–26.

[54] S. Ullman, *The Interpretation of Visual Motion* (MIT Press, Cambridge, MA, 1979).

[55] J. Weng, T. Huang and N. Ahuja, Motion and structure from two perspective views: algorithms, error analysis and error estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (5) (1989) 451–476.

[56] G. Xu, E. Nishimura and S. Tsuji, Image correspondence and segmentation by epipolar lines: theory, algorithm and applications, Technical Report, Department of Systems Engineering, Osaka University, Japan (1993).

[57] Z. Zhang, Token tracking in a cluttered scene, *Int. J. Image Vision Comput.* **12** (2) (1994) 110–120; also: Research Report No. 2072, INRIA Sophia-Antipolis (1993).

[58] Z. Zhang and O. Faugeras, Estimation of displacements from two 3D frames obtained from stereo, *IEEE Trans. Pattern Anal. Mach. Intell.* **14** (12) (1992) 1141–1156.

[59] X. Zhuang, T. Wang and P. Zhang, A highly robust estimator through partially likelihood function modeling and its application in computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **14** (1) (1992) 19–34.

[60] S. Zucker, Y. Leclerc and J. Mohammed, Continuous relaxation and local maxima selection: conditions for equivalence, *IEEE Trans. Pattern Anal. Mach. Intell.* **3** (1981) 117–127.