

# Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig

R. Deriche and Z. Zhang and Q.-T. Luong and O. Faugeras

INRIA, 2004 route des Lucioles, B.P. 93, F-06902 Sophia-Antipolis Cedex, France  
E-mail: {der,zhang,luong,faugeras}@sophia.inria.fr

**Abstract.** This paper addresses the problem of accurately and automatically recovering the epipolar geometry from an uncalibrated stereo rig and its application to the image matching problem. A robust correlation based approach that eliminates *outliers* is developed to produce a reliable set of corresponding high curvature points. These points are used to estimate the so-called *Fundamental Matrix* which is closely related to the epipolar geometry of the uncalibrated stereo rig. We show that an accurate determination of this matrix is a central problem. Using a linear criterion in the estimation of this matrix is shown to yield erroneous results. Different parametrization and non-linear criteria are then developed to take into account the specific constraints of the *Fundamental Matrix* providing more accurate results. Various experimental results on real images illustrate the approach.

## 1 Introduction

Recovering the epipolar geometry of a stereo pair of images is an important and attractive problem. It is the only geometric constraint that could be used in image matching to reduce the searching area of potential matches in the images. Until recently this step has been closely related to the calibration problem. In this paper, we are interested in recovering this information without the use of a calibration process. The proposed method consists of the following major steps: extracting points of interests, matching a set of at least 8 points, recovering as accurately as possible the so-called *Fundamental Matrix* and the epipolar geometry using a robust matching process that considers possible false matches as outliers, run for a second time the correlation process while taking into account the recovered epipolar geometry and update the *Fundamental matrix* using non-linear criteria.

The importance of the criterion to be minimized in order to correctly recover the *Fundamental Matrix* and the strong need to deal with a robust matching approach that detect and discard possible false matches, are two important points to be considered in such problem. Our work is closely related to the work of Olsen [8] and that of Shapiro and Brady [10]. Both use linear Olsen assumes that the epipolar lines are almost aligned horizontally and uses a linear method to estimate the epipolar geometry. A robust method (the M-estimator, see Sect. 3.2) is used to detect possible false matches. Shapiro and Brady also use a linear

method. The camera model is however a simplified one, namely an affine camera and correspondences are established while rejecting possible *outliers* by the use of a *regression diagnostic* processus. These two approaches (M-estimators and Regression diagnostics) work well when the percentage of outliers is small and more importantly 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 arbitrarily 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 completely be different from the valid matches. The robust technique described in this paper (i.e *Least-Median-of-Squares* ) deals with these issues and can theoretically detect as high as 50% outliers.

## 2 Geometry of Stereovision

The camera model which is most widely used is the pinhole and in the general case, the camera performs a *projective linear transformation*, rather than a mere perspective transformation. The *pixel coordinates*  $u$  and  $v$  are the only information we have if the camera is not calibrated  $\mathbf{q} = [su, sv, s]^T = \mathbf{A}\mathbf{G}\mathbf{M}$ , where  $\mathbf{M} = [X, Y, Z, 1]^T$  is the point in 3D space,  $\mathbf{A}$  is a  $3 \times 3$  transformation matrix accounting for camera sampling and optical characteristics and  $\mathbf{G}$  is a  $3 \times 4$  displacement matrix accounting for camera position and orientation. If the camera is calibrated, then  $\mathbf{A}$  is known and it is possible to use *normalized coordinates*  $\mathbf{m} = \mathbf{A}^{-1}\mathbf{q}$ , which have a direct 3D interpretation.

The epipolar geometry is the basic constraint which arises from the existence of two viewpoints. It is well-known in stereovision: for each point  $\mathbf{m}$  in the first retina, its corresponding point  $\mathbf{m}'$  lies on its epipolar line  $\mathbf{l}'_m$ .

Let us now use retinal coordinates. For a given point  $\mathbf{q}$  in the first image, the projective representation  $\mathbf{l}'_q$  of its the epipolar line in the second image is given by  $\mathbf{l}'_q = \mathbf{F}\mathbf{q}$ . Since the point  $\mathbf{q}'$  corresponding to  $\mathbf{q}$  belongs to the line  $\mathbf{l}'_q$  by definition, it follows that

$$\mathbf{q}'^T \mathbf{F}\mathbf{q} = 0 \quad (1)$$

We call the  $3 \times 3$  matrix  $\mathbf{F}$  which describes this correspondence the *fundamental matrix*. The importance of the fundamental matrix has been neglected in the literature, as almost all the work on motion has been done under the assumption that intrinsic parameters are known. In that case, the fundamental matrix reduces to an essential matrix. But if one wants to proceed only from image measurements, the fundamental matrix is the key concept, as it contains all the geometrical information relating two different images.

It can be shown that the fundamental matrix  $\mathbf{F}$  is related to the essential matrix [5]  $\mathbf{E} = \mathbf{t} \times \mathbf{R}$  by  $\mathbf{F} = \mathbf{A}^{-1T} \mathbf{E} \mathbf{A}^{-1}$ . Unlike the essential matrix, which is characterized by the two constraints described by Huang and Faugeras [3] which are the nullity of the determinant and the equality of the two non-zero singular values, the only property of the fundamental matrix is that it is of rank two. As it is also defined only up to a scale factor, the number of independent coefficients of  $\mathbf{F}$  is seven.

### 3 Matching Without the Epipolar Geometry

A slightly improved correlation based approach is used to match a set of high curvature points extracted from each image. The corner detector developed in [2], is used with a value of  $k$  equal to 0.04 for providing 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 a pixel precision. In order to recover the corner position up to a sub-pixel position, one use the model based approach we have already developed and presented in [1], 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. This step of refinement at a sub-pixel position of the high curvature points has been found important specially in the case where a small number of points are used in the minimization process performed to recover the so-called *Fundamental Matrix*.

#### 3.1 Matching by correlation

The matching process developed to put into correspondences the set of extracted points of interest is then the following one : Given a high curvature point  $\mathbf{m}_1$  in image 1, we use a correlation window of size  $(2n + 1) \times (2m + 1)$  centered around this point. We then select a rectangular search area of size  $(2d + 1) \times (2d + 1)$  around this point in the second image, and perform a correlation operation on a given window between point  $\mathbf{m}_1$  in the first image and all high curvature points  $\mathbf{m}_2$  lying within the search area in the second image. The following constraints are then applied in order to select the most consistent matches :

- **Constraint on the correlation score** : For a given couple of points to be considered as a possible set of corresponding points, the correlation score must be higher than a given threshold.
- **Constraint of the gradient direction** : The cosines of the angle between the vectors gradients of  $m_1$  and  $m_2$  is evaluated. This direction consistency measure must be greater than a given threshold for the matching to be considered consistent.
- **Constraint of curvature** This constraint is based on the magnitude of the curvature difference. The curvature sign is also used. It differentiates a convex angle from a concave one. Two points are said to be consistent in curvature if and only if the absolute value of the difference of their curvature is less than a fixed threshold.
- **Constraint of disparity** : This constraint reflects some *a priori* knowledge about the disparities between the matches points. All the feature points are supposed to have a a maximum disparity.

If the above four constraints are fulfilled, we say that the potential points considered are mutually consistent. A similarity function is computed for each consistent correspondences. This similarity function is a function of the measures defined above, weighted by the difference between the gradient norm of both points considered.

Finally before validating the match with the highest similarity value, we perform the correlation twice by reversing the roles of the two images and consider as valid only those matches for which the reverse correlation has fallen on the initial point in the first image. The two images play a symmetric role. This validity test allows us to greatly reduce the probability of error.

### 3.2 Taking into Account Possible Outliers in the Initial Correspondences

In all matches established so far, we may find two types of *outliers* due to

**bad locations.** Error in localization can be large (greater than three pixels) for some extracted points of interest. This will degrade the accuracy of the estimation.

**false matches.** False matches occurring in the correspondence process spoil the final estimate of the fundamental matrix and can render it definitely useless.

Many the so-called *robust regression* methods have been proposed that are not so easily affected by outliers [4, 9]. The reader is referred to [9, Chap. 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.

Let  $r_i$  be the *residual* of the  $i$ -th datum, i.e., the difference between the  $i$ -th 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  $\rho$  is a symmetric, positive-definite function with a unique minimum at zero. The M-estimators can be implemented as a weighted least-squares problem. This method [8, 6] reveals to be robust to outliers due to bad locations. It is, however, not robust to false matches.

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

$$\min \operatorname{med}_i r_i^2 .$$

That is, the estimates 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 locations. 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 [9, Chap. 5], as outlined below.

Given  $n$  point correspondences and their retinal image coordinates  $\tilde{\mathbf{m}}_{1i} = [u_{1i}, v_{1i}]^T$ ,  $\tilde{\mathbf{m}}_{2i} = [u_{2i}, v_{2i}]^T$  and denoting the homogeneous coordinates of  $\tilde{\mathbf{x}} = [x, y, \dots]^T$  by  $\mathbf{x}$ , i.e.,  $\mathbf{x} = [x, y, \dots, 1]^T$ , 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  $\mathbf{F}_J$ . For each  $\mathbf{F}_J$ , we can determine the median of the squared residuals, denoted by  $M_J$ , with respect to the whole set of point correspondences, i.e.,

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

We retain the estimate  $\mathbf{F}_J$  for which  $M_J$  is minimal among all  $m$   $M_J$ 's. The question now is: *How to 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  $\varepsilon$  of outliers, the probability that at least one of the  $m$  subsamples is good is given by  $P = 1 - [1 - (1 - \varepsilon)^p]^m$ . By requiring that  $P$  must be near 1, one can determine  $m$  for given values of  $p$  and  $\varepsilon$ . In our implementation, we assume  $\varepsilon = 40\%$  and require  $P = 0.99$ , thus  $m = 272$ .

As noted in [9], 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 [9, page 202] for the explanation of these magic numbers. Based on  $\hat{\sigma}$ , we can assign a weight for each correspondence:

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

where  $r_i^2 = d^2(\mathbf{m}_{2i}, \mathbf{F} \mathbf{m}_{1i}) + d^2(\mathbf{m}_{1i}, \mathbf{F}^T \mathbf{m}_{2i})$ . The correspondences having  $w_i = 0$  are outliers and should not be further taken into account. The fundamental matrix  $\mathbf{F}$  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.

## 4 Accurately Recovering the Fundamental Matrix

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 step. We have considered different linear and non-linear criteria and also considered different approaches to parametrize this matrix.

### 4.1 The linear criterion

Equation (1) can be written down as a linear and homogeneous equation in the 9 unknown elements of matrix  $\mathbf{F}$ . Given 8 or more matches we will be able, in general, to determine a unique solution for  $\mathbf{F}$ , defined up to a scale factor. For example, a singular value decomposition technique can be used.

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 numerous data points. The two main reasons for this are : The constraint  $\det(\mathbf{F}) = 0$  is not satisfied, which causes inconsistencies of the epipolar geometry near the epipoles, and the criterion is not normalized, which causes a bias in the localization of the epipoles. A detailed study of these points has been made in the technical report [7].

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  $\mathbf{F}$  is only two, and that  $\mathbf{F}$  thus depends on only 7 parameters. This could be taken into account by doing a minimization under the constraint  $\det(\mathbf{F}) = 0$ , which is a cubic polynomial in the coefficients of  $\mathbf{F}$ . However, the numerical implementations were not efficient and accurate at all. Thanks to a suggestion by Luc Robert, we can express the same constraint with an unconstrained minimization: the idea is to write one line of matrix  $\mathbf{F}$  as a linear combination of the two others, which ensures that  $\mathbf{F}$  is singular. Choosing such a representation allows us to represent  $\mathbf{F}$  by the right number of parameters, once the normalization is done.

It can be shown that  $\mathbf{F}$  can also be expressed as function of the affine coordinates  $(x, y)$  and  $(x', y')$  of the two epipoles, and the coefficients, of the homography between the two pencils of epipolar lines [7]

### 4.2 Minimizing the distance to epipolar lines

We now introduce a first non-linear approach. The idea is to use a non-linear criterion, minimizing  $\sum_i d^2(\mathbf{q}'_i, \mathbf{F}\mathbf{q}_i)$ . However, unlike the case of the linear criterion, the two images do not play a symmetric role, as the criterion determines only the epipolar lines in the second image, and should not be used to obtain the epipole in the first image. We would have to exchange the role of  $\mathbf{q}_i$  and  $\mathbf{q}'_i$  to do so. The problem with this approach is the inconsistency of the epipolar geometry between the two images. To make this more precise, if  $\mathbf{F}$  is computed by minimizing  $\sum_i d^2(\mathbf{q}'_i, \mathbf{F}\mathbf{q}_i)$  and  $\mathbf{F}'$  by minimizing  $\sum_i d^2(\mathbf{q}_i, \mathbf{F}'\mathbf{q}'_i)$ , there is no warranty that the points of the epipolar line  $\mathbf{F}\mathbf{q}$  *different from*  $\mathbf{q}'$  correspond to the points of the epipolar line  $\mathbf{F}'\mathbf{q}'$ . 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:  $\sum_i (d^2(\mathbf{q}'_i, \mathbf{F}\mathbf{q}_i) + d^2(\mathbf{q}_i, \mathbf{F}^T\mathbf{q}'_i))$ , which operates simultaneously in the two images and can be written using the expression that gives the Euclidean distance of a given point  $\mathbf{q}$  in an image to its epipolar line  $\mathbf{l}$  in the other image, and the fact that  $\mathbf{q}_i'^T \mathbf{F} \mathbf{q}_i = \mathbf{q}_i^T \mathbf{F}^T \mathbf{q}'_i$ :

$$\sum_i \left( \frac{1}{(\mathbf{F}\mathbf{q}_i)_1^2 + (\mathbf{F}\mathbf{q}_i)_2^2} + \frac{1}{(\mathbf{F}^T\mathbf{q}'_i)_1^2 + (\mathbf{F}^T\mathbf{q}'_i)_2^2} \right) (\mathbf{q}_i'^T \mathbf{F} \mathbf{q}_i)^2 \quad (2)$$

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

### 4.3 Taking into account uncertainty

Pixels are measured with some uncertainty. When using the linear criterion, we minimize a sum of terms which have different variances. It is natural to weight them so that the contribution of each of these terms to the total criterion will be inversely proportional to its variance.

Estimating the variance of each term as a function of the variance of the points  $\mathbf{q}_i$  et  $\mathbf{q}'_i$ , developping and simplifying, we obtain the following criterion :

$$\sum_i \frac{(\mathbf{q}_i'^T \mathbf{F} \mathbf{q}_i)^2}{(\mathbf{F}\mathbf{q}_i)_1^2 + (\mathbf{F}\mathbf{q}_i)_2^2 + (\mathbf{F}^T\mathbf{q}'_i)_1^2 + (\mathbf{F}^T\mathbf{q}'_i)_2^2} \quad (3)$$

Note the great similarity between this criterion and the distance criterion (2).

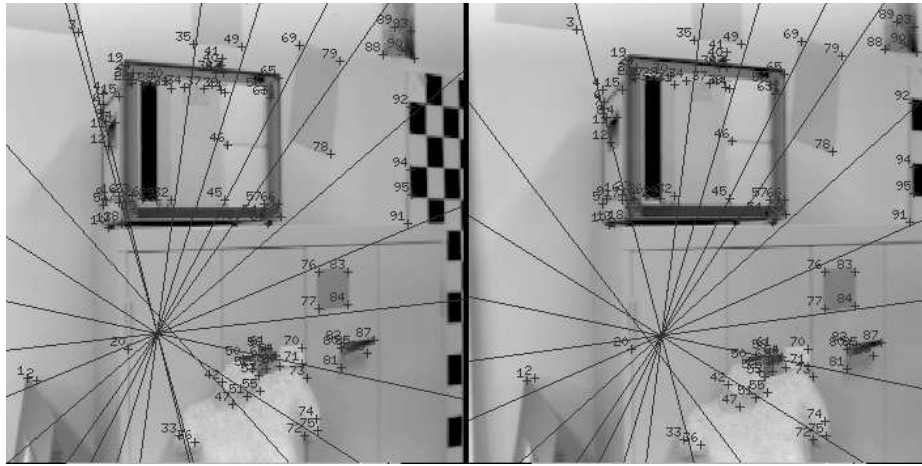
## 5 Establishing Correspondences Using the Epipolar Line Constraint and Refinement of the Fundamental Matrix

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

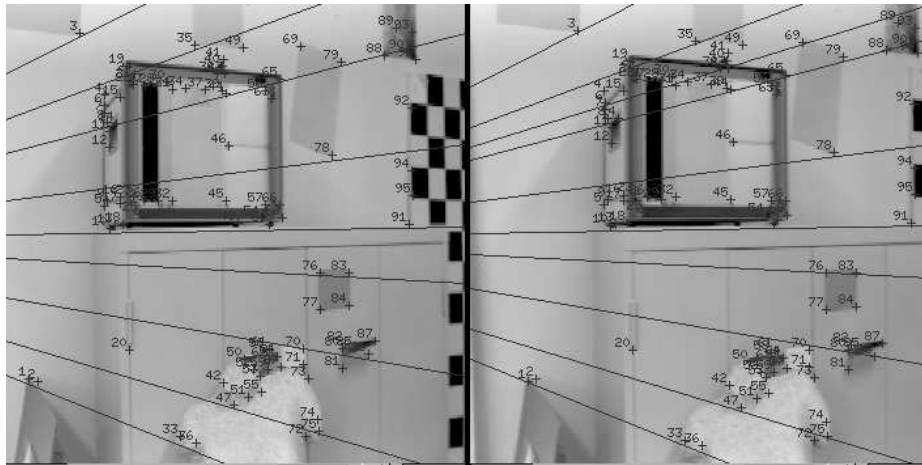
The last step of the approach is then related to the refinement of the fundamental matrix and its estimation using the larger number of correspondences produced by this matching process.

## 6 Experimental results

For the purposes of experimentation, the different types of criteria were considered and the robustness of the approach to the *outliers* has also been considered. A large number of real images were selected and an intensive experimental work has been carried out in order to test the robustness and the accuracy of the



**Fig. 1.** Indoor scene: Matched points and the epipolar geometry using the linear criterion



**Fig. 2.** Indoor scene: Matched points and the epipolar geometry recovered using the non-linear criterion

recovered epipolar geometry as well as the efficiency in detecting and rejecting *outliers*.

Figures 1 and 2 illustrate on an indoor scene the performances of the matching approach that have been proposed while using just the correlation part and without taking into account the outliers detector module. The displacement between the two images is mainly a horizontal translation toward to the left side, and we can expect that the epipoles will lie on a horizontal line passing around the middle of the image. The window size used for the correlation is  $20 \times 20$  and the parameters related to the constraints on the correlation and gradient direc-



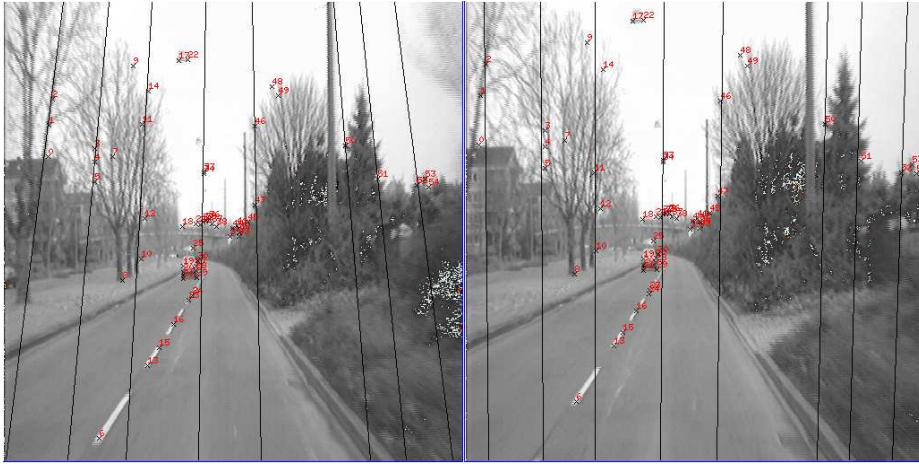


Fig. 3. Outdoor scene: Matched points by correlation and the epipolar geometry recovered using the nonlinear criterion

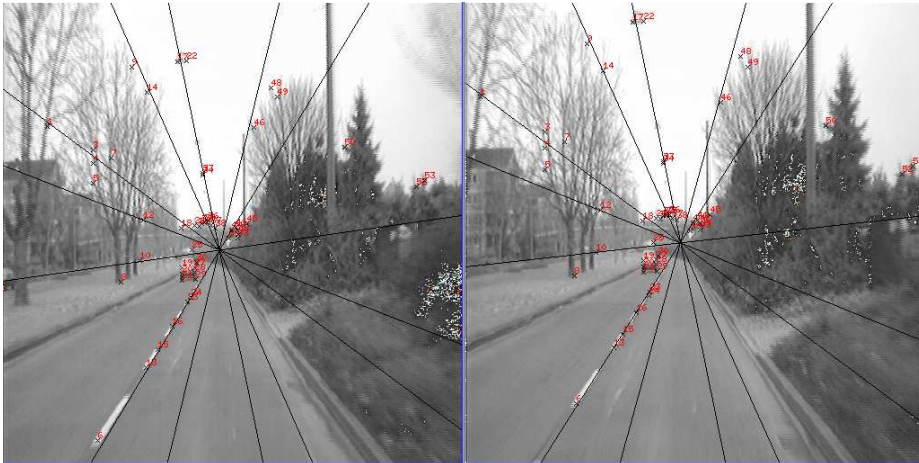


Fig. 4. Outdoor scene: Remaining matches after eliminating outliers and the corresponding epipolar geometry recovered using the nonlinear criterion

tion have been both set to .9

Figure 1 shows the set of 96 matched points using just the correlation and the associated constraints and illustrates the epipolar geometry recovered using the linear criterion. Note that the two epipolar lines that do not intersect the pencil of epipolar lines are those corresponding to the two epipoles.

Figure 2 shows the set of 96 matched points and the epipolar geometry recovered using the non-linear criterion. The positions of the epipoles are what we expected. This example illustrates the good performances of the correlation based approach provided that the deformation between the stereo pair of images

is not too large. One can also remark the robustness of the non-linear criterion compared to the linear one.

Figures 3 and 4 illustrate the necessity to take into account the possibility of having some *outliers* in the initial set of matched points. The two images have been taken by a camera mounted on a moving vehicle. It moves forward in the middle of the right lane, and the epipolar lines are thus expected to intersect to each other at a point near the middle of the right lane. Figure 3 shows the matched points recovered just by the use of the correlation technique. 55 matches have been found. One can remark that there are some points that have not been correctly matched on this image pair and one can notice that the recovered epipolar geometry is not good at all. For example, match 11 pairs two points from two different branches of the tree. Match 0 is not false visually. It describes the intersection of a tree with the roof of the house on the left side of the image. However, this point is only virtual, and it changes with the point of view. Figure 4 illustrates the performances of the *outliers* rejection module. 6 *outliers* have been detected, namely matches number 11, 0, 51, 54, 2 and 47. This fact changes significantly the recovered epipolar geometry and in particular the positions of the epipoles.

## References

- [1] R. Deriche and T. Blaszk. Recovering and Characterizing Image Features Using An Efficient Model Based Approach. In *Proceedings IEEE Conference On Computer Vision and Pattern Recognition*, New-York, June 14-17 1993.
- [2] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings Alvey Conference*, pages 189-192, 1988.
- [3] T.S. Huang and O.D. Faugeras. Some properties of the E-matrix in two view motion estimation. *IEEE Trans. PAMI*, 11:1310-1312, 1989.
- [4] P.J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.
- [5] H.C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, 293:133-135, 1981.
- [6] 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, France, December 1992.
- [7] 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.
- [8] S.I. Olsen. Epipolar line estimation. In *Proc. Second European Conf. Comput. Vision*, pages 307-311, Santa Margherita Ligure, Italy, May 1992.
- [9] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.
- [10] L. Shapiro and M. Brady. Rejecting outliers and estimating errors in an orthogonal regression framework. Tech.Report OUEL 1974/93, Dept. Engineering Science, University of Oxford, February 1993.

This article was processed using the L<sup>A</sup>T<sub>E</sub>X macro package with LLNCS style