DIIS - I3A
Universidad de Zaragoza
C/ María de Luna num. 1
E-50018 Zaragoza
Spain

Internal Report: 2003-V01

Robust line matching and estimate of homographies simultaneously[1]

Guerrero J.J., Sagüés C.

If you want to cite this report, please use the following reference instead:
Robust line matching and estimate of homographies simultaneously,
Guerrero J.J., Sagüés C., *IbPRIA, Pattern Recognition and Image Analysis*,
LNCS 2652, pages 297-307, 2003.

[1] This work was supported by projects DPI2000-1265 and DPI2000-1272.

Robust line matching and estimate of homographies simultaneously

Guerrero J.J. and Sagüés C.

Dpto. de Informática e Ingeniería de Sistemas, Univ. de Zaragoza
María de Luna 3, E-50018 ZARAGOZA (SPAIN)
email: jguerrer, csagues@posta.unizar.es *

Abstract. This paper addresses the robust matching of lines simultaneously to the computation of homographies between two views, when structure and motion are unknown. Using viewpoint non invariant measures, such as image dependent parameters, gives a lot of non matched or wrong matched features. The inclusion of projective transformations gives much better results with short computing overload. We use line features which can usually be extracted more accurately than points and they can be used in cases when there are partial occlusion. In the first stage, the lines are matched to the weighted nearest neighbor using brightness-based and geometric-based image parameters. From them, robust homographies can be computed, allowing to reject wrong matches, and growing also additional matches in the final stage. Although lines and points are dual features to compute homographies, some problems related to data representation and normalization using lines are considered. Results show that the robust technique turns out stable, and its application is useful in many situations. We have used it for robot homing and we also present automatic matching of lines at aerial images.

1 Introduction

In this paper we address the problem of robust matching of lines in two images when camera motion is unknown. Using lines instead of points has been considered by some researches [1]. Straight lines can be accurately extracted in noisy images, they capture more information than points, specially in man-made environments, and they may be used where occlusions occur.

However, line matching is more difficult than point matching because the end points of the extracted lines is not reliable. Besides that, there is not geometrical constraint, like the epipolar, for lines in two images. The putative matching of features based on image parameters has many drawbacks, giving non matched or wrong matched features. Previously the problem of wide baseline matching has been addressed establishing a viewpoint invariant affinity measure [2]. We use the homography in the matching process to select and to grow previous matches which have been obtained combining geometric and brightness image

* This work was supported by projects DPI2000-1265, DPI2000-1272.

parameters. Perspective images of plane scenes are usual in perception of man made environments, and the model to work with them is well known. Points or lines on the world plane in one image of the world plane are mapped to points or lines in the other image by a plane to plane homography [3]. This is an exact transformation for planar scenes or for small baseline image pairs. As known, there is no geometric constraint for infinite lines in two images, but the homography can be considered a first approximation for a general 3D scene.

To compute homographies, points and lines are dual geometric entities, however line-based algorithms are generally more difficult than point-based ones [4]. Thus, some particular problems related to data representation and normalization must be considered in practice. We compute homographies from corresponding lines in two images making use of classical normalization of point data [5], and avoiding singularities.

Robust estimate is actually unquestionable technique to obtain results in real situations where outliers and spurious data are present [6, 7]. In this paper the least median of squares method [8] has been used to estimate the homography. It provides not only the solution in a robust way, but also a list of previous matches that are in disagreement with it, which allows to reject wrong matches.

The simultaneous computation of matches and projective transformation between images is useful in many applications, but we have used it for robot homing. Our algorithm can also be applied in photogrammetry where points are the feature mostly used [9], but lines are plentiful in urban scenes. We have put into practice our matching with aerial images obtaining satisfactory results.

2 Basic matching

In several works, the matching is made over close images. In this context, correspondence determination by tracking geometric information along the image sequence has been proposed as a good solution [10], [11]. We determine correspondences between lines in two images of large disparity without knowledge about motion or scene structure. We use not only the geometric parameters but also the brightness attributes supplied by the contour extractor (the lines are extracted using our implementation of the method proposed by Burns [12]). So, agl and c (average grey level and contrast) of the line are combined with geometric parameters of the segments such as midpoint coordinates (x_m, y_m), the line orientation θ (in 2π range with dark on the right and bright on the left) and the length l of the extracted line.

Significant motion between views or changes on light conditions and measurements noise makes that few or none of the defined line parameters remain invariant between images.

2.1 Similarity measures

In the matching process two similarity measures are used, a geometric measure and a brightness measure. We name $\mathbf{r_g}$ the difference of geometric parameters between both images $(1, 2)$, $\mathbf{r_g} = [x_{m1} - x_{m2},\ y_{m1} - y_{m2},\ \theta_1 - \theta_2,\ l_1 - l_2]^T$.

As previously [11], we define the $\mathbf{R}$ matrix to express the uncertainty due to measurement noise in the extraction of features in each image

$$\mathbf{R} = \begin{bmatrix} \sigma_\perp^2 S^2 + \sigma_\parallel^2 C^2 & \sigma_\perp^2 CS - \sigma_\parallel^2 CS & 0 & 0 \\ \sigma_\perp^2 CS - \sigma_\parallel^2 CS & \sigma_\perp^2 C^2 + \sigma_\parallel^2 S^2 & 0 & 0 \\ 0 & 0 & 2\frac{\sigma_\perp^2}{l^2} & 0 \\ 0 & 0 & 0 & 2\sigma_\parallel^2 \end{bmatrix},$$

where $C = \cos\theta$ y $S = \sin\theta$. Location uncertainties of segment tips along the line direction and along the orthogonal direction are represented by $\sigma_\parallel$ and $\sigma_\perp$ respectively. With this geometric representation we can assume no correlation between midpoint location and θ and l parameters [10].

Additionally we define the diagonal matrix $\mathbf{P} = diag(\sigma_{x_m}^2,\ \sigma_{y_m}^2,\ \sigma_\theta^2,\ \sigma_l^2)$ to represent the uncertainty of variation of the geometric parameters due to camera motion and unknown scene structure.

Thus, from those matrixes we introduce $\mathbf{S} = \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{P}$ to weigh the variations on the geometric parameters of corresponding lines due to both, line extraction noise ($\mathbf{R}_1$ in image 1 and $\mathbf{R}_2$ in image 2) and unknown structure and motion. Note in $\mathbf{R}$ that $\sigma_\parallel$ is bigger than $\sigma_\perp$. Therefore measurement noise of x_m and y_m are coupled and the line orientation shows the direction where the measurement noise is bigger (along the line). However, in $\mathbf{P}$ the orientation does not main because the evolution of the line between images is mainly due to camera motion which is not dependent on the orientation of the image line.

The matching technique in the first stage is made to the nearest neighbor. The similarity between the parameters can be measured with a Mahalanobis distance like, $\mathbf{d_g} = \mathbf{r_g}^T \mathbf{S}^{-1} \mathbf{r_g}$.

A second similarity measure has been defined for the brightness parameters. In this case we define the diagonal matrix $\mathbf{B} = diag(\sigma_{agl}^2,\ \sigma_c^2)$, where σ_{agl} and σ_c represent the uncertainty of variations of the agl and c. Both depend on measurement noise and on changes of illumination between images.

Naming $\mathbf{r_b}$ the variation of the brightness parameters between both images, $\mathbf{r_b} = [agl_1 - agl_2,\ c_1 - c_2]^T$, the Mahalanobis distance for the similarity between the brightness parameters is, $\mathbf{d_b} = \mathbf{r_b}^T \mathbf{B}^{-1} \mathbf{r_b}$.

2.2 Matching criteria

Two image lines are stated as compatible when both, geometric and brightness variations are small. For one line in the second image to belong to the compatible set of a line in the first image, the following tests must be satisfied:

- *Geometric compatibility.* Assuming that the noise is Gaussian distributed, the similarity distance for the geometric parameters is distributed as a χ^2 with 4 d.o.f. Establishing a significance level of 5%, the compatible lines must fulfill, $\mathbf{d_g} \leq \chi_4^2(95\%)$.
- *Brightness compatibility.* Similarly, refereing to the brightness parameters, the compatible lines must fulfill, $\mathbf{d_b} \leq \chi_2^2(95\%)$.

A general Mahalanobis distance for the six parameters is not used because the correct weighting of so different information as brightness based and location based in a sole distance is difficult and could easily lead to wrong matches. Thus, compensation between high precision in some parameters with high error in other parameter is avoided.

A line in the first image can have more than one compatible line in the second image. From the compatible lines, the line having the smallest $\mathbf{d_g}$ is selected as putative match. The matching is carried out in both directions from first to second image and from second to first, in such a way that, a match $(\mathbf{n}_1, \mathbf{n}_2)$ is considered valid when the line $\mathbf{n}_2$ is the putative match of $\mathbf{n}_1$ and simultaneously $\mathbf{n}_1$ is the putative match of $\mathbf{n}_2$.

In practice the parameters $\sigma_j (j = \perp,\ \|,\ x_m,\ y_m,\ \theta,\ l,\ agl,\ c)$ introduced in $\mathbf{R}, \mathbf{P}, \mathbf{B}$ must be tuned according to the estimated image noise, expected camera motions and illumination conditions, respectively.

3 From lines to homographies

The representation of a line in the projective plane is obtained from the analytic representation of a plane through the origin: $n_1x_1 + n_2x_2 + n_3x_3 = 0$. The equation coefficients $\mathbf{n} = (n_1, n_2, n_3)^T$ correspond to the homogeneous coordinates of the projective line. All the lines written as $\lambda\mathbf{n}$ are the same than $\mathbf{n}$. The case $n_3 = 0$ corresponds to a line through the origin of the virtual image plane. As cameras have a limited field of view, observed lines have usually n_3 close to 0. Similarly, an image point $\mathbf{p} = (x, y, 1)^T$ is also an element of the projective plane and the equation $\mathbf{n} \cdot \mathbf{p} = \mathbf{n}^T \cdot \mathbf{p} = 0$ represents the belonging of point $\mathbf{p}$ to the line $\mathbf{n}$, which shows the duality of points and lines.

A projective transformation between two projective planes (1 and 2) can be represented by a linear transformation $\mathbf{T}_{21}$, in such a way that $\mathbf{p}_2 = \mathbf{T}_{21}\mathbf{p}_1$. Considering the above equations for lines in both images, we have $\mathbf{n}_2 = \left[\mathbf{T}_{21}^{-1}\right]^T \mathbf{n}_1$. A homography requires eight parameters to be completely defined, because there is an overall scale factor. A corresponding point or line gives two linear equations in terms of the elements of the homography. Thus, four corresponding lines assure a unique solution for $\mathbf{T}_{21}$, if no three of them are parallel. To have an accurate solution it is interesting to have the lines as separate in the image as possible.

3.1 Computing homographies from corresponding lines

Here, we obtain the projective transformation of points ($\mathbf{p}_2 = \mathbf{T}_{21}\mathbf{p}_1$), but using matched lines. To deduce it, we suppose the start (s) and end (e) tips of a matched line segment to be $\mathbf{p}_{s1}, \mathbf{p}_{e1}, \mathbf{p}_{s2}, \mathbf{p}_{e2}$, which usually will not be corresponding points. The line in the second image can be computed as the cross product of two of its points (in particular the observed tips) as

$$\mathbf{n}_2 = \mathbf{p}_{s2} \times \mathbf{p}_{e2} = \tilde{\mathbf{p}}_{s2}\mathbf{p}_{e2}, \tag{1}$$

where $\tilde{\mathbf{p}}_{s2}$ is the skew-symmetric matrix obtained from vector $\mathbf{p}_{s2}$.

As the tips belong to the line we have, $\mathbf{p}_{s2}^T\mathbf{n}_2 = 0$; $\mathbf{p}_{e2}^T\mathbf{n}_2 = 0$. As the tips of line in the first image once transformed also belong to the corresponding line in the second image, we can write, $\mathbf{p}_{s1}^T\mathbf{T}_{21}^T\mathbf{n}_2 = 0$; $\mathbf{p}_{e1}^T\mathbf{T}_{21}^T\mathbf{n}_2 = 0$. Combining with equation (1) we have,

$$\mathbf{p}_{s1}^T\mathbf{T}_{21}^T\tilde{\mathbf{p}}_{s2}\mathbf{p}_{e2} = 0 \; ; \; \mathbf{p}_{e1}^T\mathbf{T}_{21}^T\tilde{\mathbf{p}}_{s2}\mathbf{p}_{e2} = 0. \tag{2}$$

Therefore each couple of corresponding lines gives two homogeneous equations to compute the projective transformation, which can be determined up to a non-zero scale factor. Developing them in function of the elements of the projective transformation, we have

$$\begin{pmatrix} Ax_{s1} \; Ay_{s1} \; A \; Bx_{s1} \; By_{s1} \; B \; Cx_{s1} \; Cy_{s1} \; C \\ Ax_{e1} \; Ay_{e1} \; A \; Bx_{e1} \; By_{e1} \; B \; Cx_{e1} \; Cy_{e1} \; C \end{pmatrix} \mathbf{t} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $\mathbf{t} = (t_{11}\, t_{12}\, t_{13}\, t_{21}\, t_{22}\, t_{23}\, t_{31}\, t_{32}\, t_{33})^T$ is a vector with the elements of $\mathbf{T}_{21}$, and $A = y_{s2} - y_{e2}$, $B = x_{e2} - x_{s2}$ and $C = x_{s2}y_{e2} - x_{e2}y_{s2}$.

Using four corresponding lines, we can construct a 8×9 matrix $\mathbf{M}$. The solution corresponds with the eigenvector associated to the least eigenvalue (in this case the null eigenvalue) of the matrix $\mathbf{M}^T\,\mathbf{M}$. In order to have a reliable transformation, more than the minimum number of matches and an estimation method may be considered. Thus from n matches a $2n \times 9$ matrix $\mathbf{M}$ can be built, and the solution $\mathbf{t}$ can be obtained from SVD decomposition of this matrix [3]. In this case the relevance of each line depends on its observed length, because the cross product of the segment tips is related to the segment length.

It is known that a previous normalization of data avoids problems of numerical computation. As our formulation only uses image coordinates of observed tips, data normalization proposed for points [5] has been used.

3.2 Robust estimation

The least squares method assumes that all the measures can be interpreted with the same model, which makes it to be very sensitive to out of norm data. Robust estimation tries to avoid the outliers in the computation of the estimate. From the existing robust estimation methods [6], we have chosen the least median of squares method. This method makes a search in the space of solutions obtained from subsets of minimum number of matches. The algorithm to obtain an estimate with this method can be summarized as follows:

1. A Monte-Carlo technique is used to randomly select m subsets of 4 features.
2. For each subset S, we compute a solution in closed form $\mathbf{T}_S$.
3. For each solution $\mathbf{T}_S$, the median M_S of the squares of the residue with respect to all the matches is computed.
4. We store the solution $\mathbf{T}_S$ which gives the minimum median M_S.

A selection of m subsets is good if at least in one subset the 4 matches are good. Assuming a ratio ϵ of outliers, the probability of one of them been good can be obtained [8] as, $P = 1 - \left[1 - (1-\epsilon)^4\right]^m$. For example, if we want a probability $P = 0.999$ of having one good at least, with $\epsilon = 35\%$ of outliers, the number of subsets m should be 34.

Once the solution has been obtained, the outliers can be selected from those of maximum residue. As in [6] the threshold is fitted proportional to the standard deviation of the residue, estimated as [8], $\hat{\sigma} = 1.4826\,[1 + 5/(n-4)]\,\sqrt{M_S}$. Assuming that the measurement error is Gaussian with zero mean and standard deviation σ, then the square of the residues follows a χ^2_2 distribution with 2 degrees of freedom. Taking, for example, that 95% probability is established for the line to fit in the homography (inlier) then the threshold will be fixed to $5.99\,\hat{\sigma}^2$.

4 Final matches

From here on, we introduce the geometrical constraint introduced by the estimated homograpy to get a bigger set of matches. Actually we compute an only homography in the image. This would be right if the scene points were on a plane. Although this fails in some situations, the results are good when the distance from the camera to the scene is large enough with respect to the baseline. For example, this assumption gives very good results in robot homing, where image disparity is mainly due to camera rotation, and therefore the sole homography captures the robot orientation, that is the most useful information for a robot to correct its trajectory. We have also made some experiments to segment into several scene planes, to obtain line matching in more general situations. This segmentation of planes could be very useful to make automatic 3D model of urban scenes.

Our objective here is to obtain at the end of the process more good matches, also eliminating wrong matches given by the basic matching. Thus final matches are composed by two sets. The first one is obtained from the matches selected after the robust computation of the homography that passe additionally an overlapping test compatible with the transformation of the segment tips. The second set of matches is obtained taking all the segments not matched initially and those being rejected previously. With this set of lines a matching process similar to the basic matching is carried out. However, now the matching is made to the nearest neighbor segment transformed with the homography. The transformation is applied to the end tips of the image segments using the homography $\mathbf{T}_{21}$ to find, not only compatible lines but also compatible segments in the same line.

In the first stage of the matching process there was no previous knowing of camera motion. However in this second step the computed homography provides information about expected disparity and therefore the uncertainty of geometric variations can be reduced. A new tuning of σ_{x_m}, σ_{y_m}, σ_θ and σ_l, must be considered. To automate the process, a global reduction of these parameters has been proposed and tested in several situations, obtaining good results with reductions about 1/5. As the measurement noise ($\sigma_\parallel$ and $\sigma_\perp$) has not changed, the initial

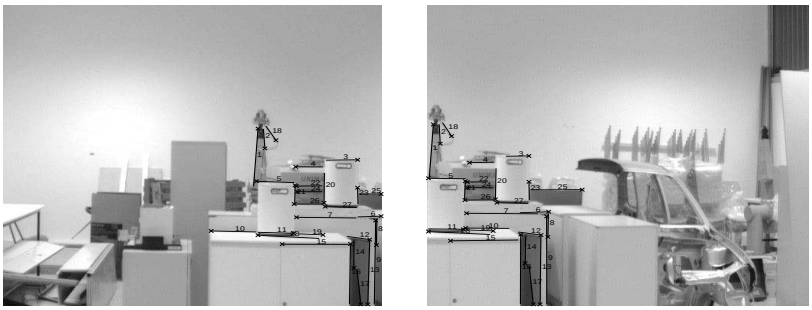

Fig. 1. Images showing the final matches when the robot rotates 18 degrees. Only one match is no good (10), which is a wrong match as segment although good as line.

tuning is maintained in this second step. Note that the brightness compatibility set is the initially computed, and therefore it must not be repeated.

5 Experimental Results

A set of experiments with different kind of images has been carried out to test the algorithm proposed. The images correspond to different applications: Indoor robot homing, architectural models and aerial images. In the algorithms there are extraction parameters which allows to obtain more or less lines according to its minimum length and minimum gradient. There are also parameters to match the lines, whose tuning has turned out simple and quite intuitive. In the experiments we have used some small variations with respect to the following tentative tuning parameters $\sigma_\perp = 1$, $\sigma_\parallel = 10$, $\sigma_{agl} = 8$, $\sigma_c = 4$, $\sigma_{xm} = 60$, $\sigma_{ym} = 20$, $\sigma_\theta = 2$, $\sigma_l = 10$. When these changes are important we indicate them in the particular experiment.

We have applied the algorithm presented for robot homing. In this application the robot must go to previously learnt positions using a camera [13]. The robot corrects its heading from the computed projective transformation between learnt and actual images.

In this experiment a set of robot rotations (from 2 to 20 degrees) has been made. The camera center is about 30 $cm.$ out of the axis of rotation of the robot and therefore this camera motion has a short baseline. In Table 1 the number of matches in the three steps with this set of camera motions are shown. The number of lines extracted in the reference image is 112.

From this experiment the progressive advantage of the simultaneous computation of the homography and matching can be seen. When the image disparity is small, the robust estimation of the homography does not improve the basic matching. However, with a disparity close to 70% of the image size, the basic matching produces a high ratio of wrong matches ($> 30\%$), that are automatically corrected in the final matching. We observe that in this case the system also works even with a large image disparity.

Robot Rotation	σ_{x_m}	Basic	After_$\mathbf{T}_{21}$	Final
4°	60	73 (5W)	56 (1W)	76 (0W)
8°	60	53 (6W)	31 (0W)	52 (0W)
12°	100	41 (9W)	30 (2W)	33 (1W)
16°	100	27 (9W)	17 (3W)	30 (1W)
20°	140	28 (10W)	17 (1W)	24 (0W)

Table 1. Number of matches in the tree steps of the algorithm, with some robot rotation, indicating also the number of wrong matches (W). Here, the matches that are good as lines but wrong as segments (not overlapped) are considered wrong.

Fig. 2. Images of other indoor scene showing the final matches (66). All of them are good when considered as lines, although two of them are wrong as segments.

To simplify, only the images corresponding to the 18 degrees of robot rotation are shown. A 38 % of wrong matches are given by the basic matching. At the final matching stage, all the matches are good when considered as lines, although one of them can be considered wrong as segment (Fig. 1).

Other experiments have been carried out indoor. In Fig. 2 we show the two images taken with a stereo system having $30cm.$ of baseline. The number of lines extracted are 83 and 93 respectively. The basic matching gives a 45 matches but 16% are wrong matches. After the computation of the homography all are good but only 35 matches remain. At the final stage 66 matches are given, and only two can been considered wrong as segment, although they are good as lines.

5.1 Aerial images

In this experiment, two aerial images with quite large stereo between them are used (Fig. 3). In photogrammetry applications putative matching has usually a high ratio of spurious results. This is confirmed in our case, where the basic matching has given a ratio of wrong matches higher than 50% , which is the theoretical limit of least median of squares method. However, if we select a smaller percent of the squares of the residue instead of the median, the robust method works properly. The results in Fig. 3 have been obtained with a percent of 30%. The basic matching provides 121 matches, 64 of them being wrong.

The robust computation of the homography provides 55 matches, 11 of them being wrong as segment but good as infinite line. Among the 105 final matches, there are only 3 wrong matches which correspond to contiguous cars. Note that the final matches are duplicated with respect to the matches obtained with the homography. Note also that the final matches selected are mainly located on the ground. There are some lines on the roofs of the buildings but they are nearly parallel to the flight of the camera which is coherent with the model of homography used.

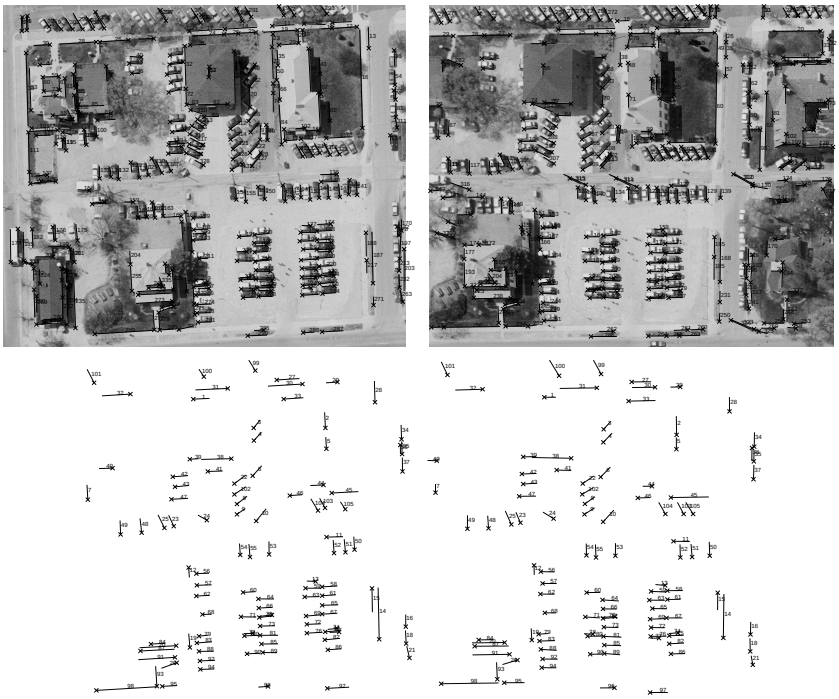

Fig. 3. Two aerial images with quite large stereo. The first row shows the lines extracted (approximately 300 lines/image). From them, the basic matching provides 121 matches (64 being wrong). Second row shows the matches at the final stage (105 matches, 3 being wrong that are corresponding to contiguous cars).

6 Conclusions

We have presented and tested a method to automatically obtain matches of lines simultaneously to the robust computation of homographies. The robust computation works especially well to eliminate outliers which may appear when matching is based on image properties and there is no information of scene structure

or camera motion. The homographies are computed from lines extracted and the use of lines has advantages with respect to the use of points. The geometric mapping between uncalibrated images provided by the homography turns out useful to grow matches and to eliminate wrong matches.

All the work is made automatically with only some previous tuning of parameters related to expected camera motion. As can be seen in the experiments, the proposed algorithm works with different types of scenes and the tuning phase is simple and intuitive. As limitation of this work, the matching depends on the mapping between the lines and the homography computed. So, plane scenes or situations where disparity is mainly due to rotation, give the best results. However, it is also possible to compute several homographies according to scene structure which is the goal we are actually working for.

References

1. Schmid, C., Zisserman, A.: Automatic line maching across views. In: IEEE Conference on CVPR. (1997) 666–671
2. Pritchett, P., Zisserman, A.: Wide baseline stereo matching. In: IEEE Conference on Computer Vision. (1998) 754–760
3. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
4. Quan, L., Kanade, T.: Affine structure from line correspondences with uncalibrated affine cameras. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 834–845
5. Hartley, R.: In defense of the eight-point algorithm. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 580–593
6. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. Rapport de recherche RR-2676, I.N.R.I.A., Sophia-Antipolis, France (1995)
7. Torr, P., Murray, D.: The development and comparison of robust methods for estimating the fundamental matrix. International Journal of Computer Vision **24** (1997) 271–300
8. Rousseeuw, P., Leroy, A.: Robust Regression and Outlier Detection. John Wiley, New York (1987)
9. Habib, A., Kelley, D.: Automatic relative orientation of large scale imagery over urban areas using modified iterated hough transform. Journal of Photogrammetry and Remote Sensing **56** (2001) 29–41
10. Deriche, R., Faugeras, O.: Tracking line segments. In: First European Conference on Computer Vision, Antibes, France (1990) 259–268
11. Guerrero, J., Sagüés, C.: Tracking features with camera maneuvering for vision-based navigation. Journal of Robotic Systems **15** (1998) 191–206
12. Burns, J., Hanson, A., Riseman, E.: Extracting straight lines. IEEE Trans. on Pattern Analysis and Machine Intelligence **8** (1986) 425–455
13. Guerrero, J., Sagüés, C.: Robot homing from homographies. Technical report, D.I.I.S. Universidad de Zaragoza (2002)