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

Internal Report: 2005-V15

Robust line matching in image pairs of scenes with dominant planes[1]

C. Sagüés, J.J. Guerrero

If you want to cite this report, please use the following reference instead:
Robust line matching in image pairs of scenes with dominant planes, C. Sagüés, J.J. Guerrero, *Optical Engineering*, Vol. 45, no. 6, pp. 06724 1-12, 2006.

[1]This work was supported by projects DPI2000-1272 and DPI2003-07986.

Robust line matching in image pairs of scenes with dominant planes

Sagüés C., Guerrero J.J.

DIIS - I3A, Universidad de Zaragoza
C/ María de Luna num. 1
E-50018 Zaragoza, Spain.
Phone 34-976-762349, Fax 34-976-761914.
csagues, jguerrer@unizar.es

Abstract

We address the robust matching of lines between two views, when camera motion is unknown and dominant planar structures are viewed. The use of viewpoint non invariant measures 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 where there are partial occlusions. In the first stage, the lines are matched to the weighted nearest neighbor using brightness and geometric-based image parameters. From them, robust homographies can be computed, which allows to reject wrong matches and to add new good matches. When two or more planes are observed, the corresponding homographies can be computed and they can be used to obtain also the fundamental matrix, which gives constraints for the whole scene. The simultaneous computation of matches and projective transformations is extremely useful in many applications. It can be seen that the proposal works in different situations requiring only a simple and intuitive parameter tuning.

Keywords: Machine vision, matching, lines, homographies, uncalibrated vision, plane segmentation, multi-plane scene.

1 Introduction

We address the problem of robust matching of lines in two images when the camera motion is unknown. Using lines instead of points has been considered in other works [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 when occlusions occur. The invariant regions used to match images, as the well known SIFT [2], often fails due to lack of textures in man made environments with homogeneous surfaces [3]. Fortunately in this kind of scenes line segments are available to match them. 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 nonmatched or wrong matched features.

Previously, the problem of wide baseline matching has been addressed establishing a viewpoint invariant affinity measure [4]. We use the homography in the matching process to select and to grow previous matches which have been obtained combining geometric and brightness image parameters. Scenes with dominant planes are usual in man made environments, and the model to work with multiple views of them is well known. Points or lines on the world plane projected in one image are mapped to the corresponding points or lines in the other image by a plane-to-plane homography [5]. As known, there is no geometric constraint for infinite lines in two images of a general scene, but the homography is a good constraint for scenes with dominant planes or small baseline image pairs, although they have large relative rotation.

Robust estimation techniques are currently unquestionable to obtain results in real situations where outliers and spurious data are present [6, 7]. In this work the least median of squares and the Ransac methods have been used to estimate the homography. They provide 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. To compute the homographies, points and lines are dual geometric entities, however line-based algorithms are generally less usual than point-based ones [8]. Thus, some particular problems related to data representation and normalization must be considered in practice. We compute the homographies from corresponding lines in two images making use of classical normalization of point data [9].

The simultaneous computation of matches and projective transformation between images is useful in many applications and the search of one homography is right if the scene features are on a plane or for small baseline image pairs. Although this transformation is not exact in general situations, the practical results are also 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 [10], where image disparity is mainly due to camera rotation, and therefore a sole homography captures the robot orientation, that is the most useful information for a robot to correct its trajectory.

Our proposal can also be applied in photogrammetry for automatic relative orientation of convergent image pairs. In this context, points are the feature mostly used [11], but lines are plentiful in urban scenes. We have put into practice our proposal with aerial images and images of buildings, obtaining satisfactory results.

If the lines are in several planes, the corresponding homographies can be iteratively obtained. From them, the fundamental matrix, which gives a general constraint for the whole scene, can be computed [12]. We have also made experiments to segment several scene planes when they exist, obtaining line matching in that situation. This segmentation of planes could be very useful to make automatically 3D model of urban scenes.

After this introduction, we present in §2 the matching of lines using geometric and brightness parameters. The robust estimation of homographies from lines is explained in §3. After that, we present in §4 the process to obtain the final matches using the geometrical constraint given by the homography. Then we present the multi-plane method §5, and the useful case of having only vertical features §6. Experimental results with real images are presented in §7. Finally, §8 is devoted to expose the conclusions. A previous version of this matcher was presented in [13], but now we have completed and extended the work showing the multi plane method and the special case of having only vertical lines. We also include experiments made with the new software version which implements different robust

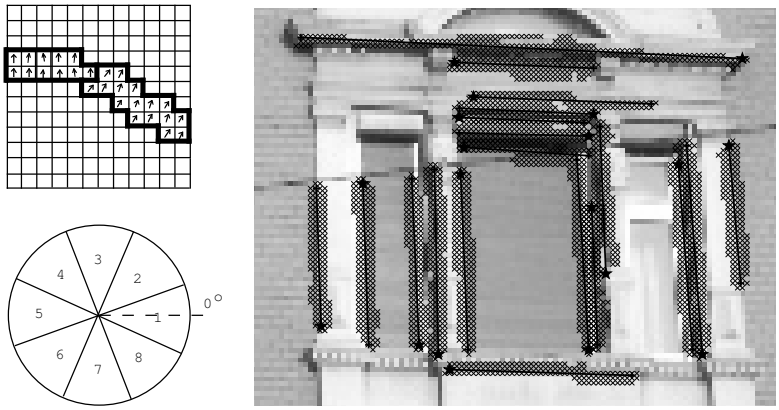

Figure 1: Left: Segmentation of the image into regions which have similar orientation of brightness gradient, LSR. Right: Zoom of the LSRs extracted on real image. From the LSR, we obtain the line with its geometric description and also attributes related to its brightness and contrast used for the line matching.

techniques.

2 Basic matching

As said, the initial matching of straight lines in the images is made to the weighted nearest neighbor. These lines are extracted using our implementation of the method proposed by Burns [14], which computes spatial brightness gradients to detect the lines in the image. Pixels having the gradient magnitude larger than a threshold are grouped into regions of similar direction of brightness gradient. These groups are named Line Support Regions (LSR). A least-squares fitting into the LSR is used to obtain the line, and the detector gives not only the geometrical parameters of the lines, but also several attributes related to their brightness and some quality attributes, that provide very useful information to select and identify the lines (Fig. 1).

We use not only the geometric parameters, but also the brightness attributes supplied by the contour extractor. So, agl and c (average grey level and contrast) in the LSR, 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 its length l.

In several works, the matching is made over close images. In this field, correspondence determination by tracking geometric information along the image sequence has been proposed as a good solution [15]. We determine correspondences between lines in two images of large disparity in the uncalibrated case. Significant motion between views or changes on light conditions of the viewed scenes makes that few or none of the defined line parameters remain invariant between images. So, for example, approaching motions to the objects gives bigger lines which are also going out the image. Besides that, measurement noise introduces more problems to the desired invariance of parameters.

Now, it must be stated that a line segment is considered to select the nearest neighbor, but assuming higher uncertainty along the line than across it. However, to compute afterwards projective transformations between images, only the infinite line information is considered. This provides an accurate solution being robust to partial occlusions without assuming that

the tips of the extracted line are the same.

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 $(1, 2)$

$$\mathbf{r_g} = \begin{bmatrix} x_{m1} - x_{m2} \\ y_{m1} - y_{m2} \\ \theta_1 - \theta_2 \\ l_1 - l_2 \end{bmatrix}.$$

As previously [15], 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.

Additionally we define the $\mathbf{P}$ matrix to represent the uncertainty on the difference of the geometric parameters due to camera motion and unknown scene structure

$$\mathbf{P} = \begin{bmatrix} \sigma_{x_m}^2 & 0 & 0 & 0 \\ 0 & \sigma_{y_m}^2 & 0 & 0 \\ 0 & 0 & \sigma_\theta^2 & 0 \\ 0 & 0 & 0 & \sigma_l^2 \end{bmatrix},$$

where $\sigma_i, (i = x_m,\ y_m,\ \theta,\ l)$ represents the uncertainty of variation of the geometric parameters of the image segment.

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 the first image and $\mathbf{R}_2$ in the second image) 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 matter because the evolution of the line between images is mainly due to camera motion, which is not dependent on the line orientation in the image.

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 as

$$\mathbf{d_g} = \mathbf{r_g}^T \mathbf{S}^{-1} \mathbf{r_g}.$$

The second similarity measure has been defined for the brightness parameters. In this case we have

$$\mathbf{B} = \begin{bmatrix} \sigma_{agl}^2 & 0 \\ 0 & \sigma_c^2 \end{bmatrix},$$

where σ_{agl} and σ_c represent the uncertainty of variations of the average gray level and the contrast of the line. Both depend on measurement noises and on changes of illumination between the images.

Naming $\mathbf{r_b}$ the variation of the brightness parameters between both images

$$\mathbf{r_b} = \begin{bmatrix} agl_1 - agl_2 \\ c_1 - c_2 \end{bmatrix},$$

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

  Under assumption of Gaussian noise, the similarity distance for the geometric parameters is distributed as a χ^2 with 4 degrees of freedom. Establishing a significance level of 5%, the compatible lines must fulfill,

  $$\mathbf{d_g} \leq \chi^2_4(95\%).$$

- Brightness compatibility

  Similarly referring 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 parameters 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 the first to second image and from second to first. 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 $\sigma_j (j = \perp,\ \parallel,\ x_m,\ y_m,\ \theta,\ l,\ agl,\ c)$ introduced in $\mathbf{R}, \mathbf{P}, \mathbf{B}$ must be tuned according to the estimated image noise, expected camera motion 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 coefficients of $\mathbf{n} = (n_1, n_2, n_3)^T$ correspond to the homogeneous coordinates of the projective line. All the lines written as $\lambda\mathbf{n}$ are the same than $\mathbf{n}$. Similarly, an image point $\mathbf{p} = (x, y, 1)^T$ is also an element of the projective plane and the equation $\mathbf{p}^T \cdot \mathbf{n} = \mathbf{n}^T \cdot \mathbf{p} = 0$ represents that the point $\mathbf{p}$ is on 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 $\mathbf{H}_{21}$, in such a way that $\mathbf{p}_2 = \mathbf{H}_{21}\mathbf{p}_1$. Considering the above equations for lines in both images, we have $\mathbf{n}_2 = \left[\mathbf{H}_{21}^{-1}\right]^T \mathbf{n}_1$. This homography requires eight parameters to be completely defined, because it is defined up to an overall scale factor. A pair of corresponding points or lines gives two linear equations in terms of the elements of the homography. Thus, four pairs of corresponding lines assure a unique solution for $\mathbf{H}_{21}$, if no three of them are parallel.

3.1 Computing homographies from corresponding lines

Here, we obtain the projective transformation of points ($\mathbf{p}_2 = \mathbf{H}_{21}\mathbf{p}_1$), 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 the line in the first image once transformed also belong to the corresponding line in the second image, we can write, $\mathbf{p}_{s1}^T\mathbf{H}_{21}^T\mathbf{n}_2 = 0$; $\mathbf{p}_{e1}^T\mathbf{H}_{21}^T\mathbf{n}_2 = 0$. Combining with equation (1) it turns out,

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

Therefore each pair 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

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

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

Using four line correspondences, we can construct a 8×9 matrix $\mathbf{M}$. The solution for $\mathbf{h}$ is the eigenvector associated to the least eigenvalue (in this case the null eigenvalue) of the matrix $\mathbf{M}^T\,\mathbf{M}$. If we have more than the minimum number of good matches, an estimation method may be considered to obtain a better solution. Thus n matches give a $2n \times 9$ matrix $\mathbf{M}$, and the solution $\mathbf{h}$ can be obtained using the singular value decomposition of this matrix [5].

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 [9] has been used. In our case the normalization makes the x and y coordinates to be centered in an image which has unitary width and height.

Equations (2) can be used as a residue to designate inliers and outliers with respect to a candidate homography. In this case the residue is an algebraic distance and the relevance of each line depends on its observed length, because the cross product of the segment tips is related to its length.

Alternatively a geometric distance in the image can be used. It is obtained from the sum of the squared distances between the segment tips and the corresponding transformed line

$$d^2 = \frac{(\mathbf{p}_{s1}^T \mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})^2}{(\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_x^2 + (\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_y^2} + \frac{(\mathbf{p}_{e1}^T \mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})^2}{(\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_x^2 + (\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_y^2}, \tag{3}$$

where subindexes $()_x$ and $()_y$ indicate first and second component of the vector respectively.

Doing the same for both images and with both observed segments, the squared distance with geometric meaning is:

$$d^2 = \frac{(\mathbf{p}_{s1}^T \mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})^2 + (\mathbf{p}_{e1}^T \mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})^2}{(\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_x^2 + (\mathbf{H}_{21}^T \tilde{\mathbf{p}}_{s2} \mathbf{p}_{e2})_y^2} + \frac{(\mathbf{p}_{s2}^T \mathbf{H}_{21}^{-T} \tilde{\mathbf{p}}_{s1} \mathbf{p}_{e1})^2 + (\mathbf{p}_{e2}^T \mathbf{H}_{21}^{-T} \tilde{\mathbf{p}}_{s1} \mathbf{p}_{e1})^2}{(\mathbf{H}_{21}^{-T} \tilde{\mathbf{p}}_{s1} \mathbf{p}_{e1})_x^2 + (\mathbf{H}_{21}^{-T} \tilde{\mathbf{p}}_{s1} \mathbf{p}_{e1})_y^2} \tag{4}$$

3.2 Robust estimation

The least squares method assumes that all the measures can be interpreted with the same model, which makes it very sensitive to outliers. Robust estimation tries to avoid the outliers in the computation of the estimate.

From the existing robust estimation methods [6], we have initially 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. If we need a minimum of 4 matches to compute the projective transformation, and there are a total of n matches, then the search space will be obtained from the combinations of n elements taken 4 by 4. As that is computationally too expensive, several subsets of 4 matches are randomly chosen. 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 $\mathbf{H}_S$.
3. For each solution $\mathbf{H}_S$, the median M_S of the squares of the residue with respect to all the matches is computed.
4. We store the solution $\mathbf{H}_S$ which gives the minimum median M_S.

A selection of m subsets is good if at least one of them has no outliers. Assuming a ratio ϵ of outliers, the probability of one subset to be good can be obtained [16] as, $P = 1 - [1 - (1 - \epsilon)^4]^m$. Thus assuming a probability $(1 - P)$ of mismatch the computation, the number of subset m to consider can be obtained. For example, if we want a probability $P = 0.99$ of one of them being good, having $\epsilon = 35\%$ of outliers, the number of subsets m

should be 24. Having $\epsilon = 70\%$ of outliers the number of subsets m should be 566 and a minor quantile instead of the median must be selected. Anyway, a reduction of ten times in the probability of mismatch the computation ($P = 0.999$) only yields a computational cost of 0.5 times larger ($m = 36$ with $\epsilon = 35\%$ and $m = 849$ with $\epsilon = 70\%$).

3.3 Rejecting wrong basic matches

Once the solution has been obtained, the outliers can be selected from those of higher residue. Good matches can be selected between those of residue smaller than a threshold. As in [6] the threshold is fitted proportional to the standard deviation of the error that can be estimated [16] as,

$$\hat{\sigma} = 1.4826\,[1 + 5/(n-4)]\,\sqrt{M_S}.$$

Assuming that the measurement error for in inliers is Gaussian with zero mean and standard deviation σ, then the square of the residues is a sum of Gaussian variables and follows a χ^2_2 distribution with 2 degrees of freedom. Taking, for example, a 5% of probability of rejecting a line being correct, the threshold will be fixed to $5.99\,\hat{\sigma}^2$.

4 Final matches

From here on, we introduce the geometrical constraint imposed by the estimated homography to get a bigger set of matches. 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 passing 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 $\mathbf{H}_{21}$ 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 knowledge of the 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. So, a new tuning of σ_{x_m}, σ_{y_m}, σ_θ and σ_l, is considered. To automate the process, a global reduction of these parameters has been proposed and tested in several situations, obtaining good results with reductions of about 1/5.

As the measurement noise ($\sigma_\parallel$ and $\sigma_\perp$) has not changed, the initial tuning of these parameters is maintained in this second step. Note that the brightness compatibility set is the initially computed, and therefore it is not repeated.

5 Different planes

When lines are in different planes, the corresponding homographies can iteratively be computed. Previously, we have explained how to determine a homography assuming that some

of the lines are outliers. The first homography can be computed in this way assuming that a certain percentage of matched lines are good matches and belong to a plane, and the other matches are bad or they cannot be explained by this homography. We do not have a priori knowledge about which plane of the scene is going to be extracted the first, but the probability of a plane being chosen increases with the number of lines in it. The most reasonable way to extract a second plane is eliminating the lines used to determine the first homography. Here we are assuming that lines belonging to the first plane do not belong to the second, which is true except for the intersection line between both planes. This idea of eliminating lines already used will apply if more planes are extracted.

On the other hand, in the multi-plane case the percentage of outliers for each homography will be high even all the matches are good. As the least median of squares method has a breakdown point in 50% of outliers, we have used in this case the Ransac method [17] which needs a tuning threshold, but works with a less demanding breakdown point.

Besides that, all the homographies obtained from an image pair are related between them. This relation can be expressed with the fundamental matrix ($\mathbf{F}_{21}$) and the epipole ($\mathbf{e}_2$), which is unique for an image pair, through the following equation:

$$\mathbf{F}_{21} = [\mathbf{e}_2]_\times \mathbf{H}_{21}^{\pi}, \tag{5}$$

$[\mathbf{e}_2]_\times$ being the skew matrix corresponding to epipole $\mathbf{e}_2$ of the second image and $\mathbf{H}_{21}^{\pi}$ the homography between first and second image through plane π. If at least two homographies $(\mathbf{H}_{21}^{\pi_1}, \mathbf{H}_{21}^{\pi_2})$ can be computed between both images corresponding to two planes (π_1, π_2), an homology $\mathbf{H} = \mathbf{H}_{21}^{\pi_1} \cdot (\mathbf{H}_{21}^{\pi_2})^{-1}$, that is a mapping from one image into itself, exists.

It turns out that the homology has two equal eigenvalues [5]. The third one is related to the motion and to the structure of the scene. These eigenvalues can be used to test when two different planes have been computed, and then the epipole and the intersection of the planes can be also obtained. The epipole is the eigenvector corresponding to the non-unary eigenvalue of the homology and the other two eigenvectors define the intersection line of the planes [18]. In case of small baseline or if there exist only one plane in the scene, the epipolar geometry is not defined and only one homography can be computed. So possible homology $\mathbf{H}$ will be close to identity, up to scale.

In practice, we propose a filter to notice the goodness of the homology using these ideas. Firstly, we normalize the homology dividing by the median eigenvalue. If there are no two unary eigenvalues, up to a threshold, then the computation of the second homography is repeated. If the three eigenvalues are similar we search if the homology is close to identity. In this case two similar homographies would explain the scene or the motion, and the fundamental matrix cannot be computed. In other case, two planes and two coherent homographies are given as result of the process. From them the fundamental matrix can be computed [12].

6 Particular case: Vertical lines

A special case of interest in some applications can be considered reducing in one dimension the problem. For example, when robot moves at man made environments, the motion is on a plane and vertical lines give enough information to carry out visual control. For vertical lines, only the x coordinate is relevant to compute the homography, therefore the problem

is simplified. The homography now has three parameters and each vertical line gives one equation, therefore three vertical line correspondences are enough.

The matching is quite similar to the presented for the lines in all directions. In this case the parameter referred to the orientation of the lines is discarded because it has no sense, although the lines are grouped in two incompatible subsets, those having 90 degrees of orientation and those with 270 degrees of orientation.

6.1 Estimation of the homography

In this case, as the dimension of the problem is reduced, only the x_m coordinate is relevant for the computation of the projective transformation $\mathbf{H}_{21}$. We can use the point based formulation which is simpler because a vertical line can be projectively represented as a point with $(x_m, 1)$ coordinates. Then, we can write, up to a scale factor

$$\begin{pmatrix} \lambda\, x_{m2} \\ \lambda \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_{m1} \\ 1 \end{pmatrix}$$

Therefore each pair of corresponding vertical lines having x_{m1} and x_{m2} coordinates respectively provides one equation to solve the matrix $\mathbf{H}_{21}$

$$\begin{pmatrix} x_{m1} & 1 & -x_{m1}x_{m2} & -x_{m2} \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \end{pmatrix} = 0.$$

With the coordinates of at least $n = 3$ vertical line correspondences we can construct a $n \times 4$ matrix $\mathbf{M}$. The homography solution is the eigenvector associated to the least eigenvalue of the $\mathbf{M}^T\mathbf{M}$ matrix.

As above before, we use a robust method to compute the projective transformation. With vertical lines the problem is computationally simpler. For example with $P = 0.99$ and 70% of outliers only 168 sets of 3 matches must be considered (in the general case we need 566 sets of 4 matches). As told before, the estimated homography is also used to obtain a better set of matched lines.

7 Experimental Results

A set of experiments with different kind of images has been carried out to test the proposal. The images correspond to different applications: Indoor robot homing, images of buildings, and aerial images.

In the algorithms there are extraction parameters to obtain more or less line segments 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 the following tuning parameters $\sigma_\perp = 1,\ \sigma_\parallel = 10,\ \sigma_{agl} = 8,\ \sigma_c = 4,\ \sigma_{xm} = 60,\ \sigma_{ym} = 20,\ \sigma_\theta = 2,\ \sigma_l = 10$, or small variation with respect to them depending on expected image disparity.

We have applied the algorithm presented for robot homing. In this application the robot must go to previously learnt positions using a camera [10]. The robot corrects its heading from the computed projective transformation between target and current images. In this

Robot Rotation	σ_{x_m}	Basic	After_$\mathbf{H}_{21}$	Final
2°	60	92 (1W)	78 (0W)	90 (0W)
4°	60	73 (5W)	56 (1W)	76 (0W)
6°	60	63 (4W)	47 (1W)	63 (0W)
8°	60	53 (6W)	31 (0W)	52 (0W)
10°	60	47 (9W)	35 (1W)	50 (0W)
12°	100	41 (9W)	30 (2W)	33 (1W)
14°	100	36 (12W)	24 (2W)	32 (1W)
16°	100	27 (9W)	17 (3W)	30 (1W)
18°	140	37 (14W)	24 (3W)	28 (1W)
20°	140	28 (10W)	17 (1W)	24 (0W)

Table 1: Number of matches with several robot rotation, indicating also the number of wrong matches (W). First column show the robot rotation. Second shows the tuning of σ_{x_m} used in the matching. Third shows the basic matches. Forth shows the matches obtained after the constraint imposed by the homography and fifth column shows the final matches. Here, the matches that are good as lines but wrong as segments (they are not overlapped) are counted as wrong matches.

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 the table 1 the number of matches in the successive steps with this set of camera motions are shown. The number of lines extracted in the reference image is 112. Here, 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 between two images having a large image disparity.

To simplify, only the images corresponding to the 18 degrees of robot rotation are shown (Fig. 2). 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.

We have also used two aerial images with large baseline (Fig. 3). In photogrammetry applications usually putative matching has 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 quantile of the squares of the residue instead of the median, the robust method works properly. The results have been obtained with the Least 30-Quantile of Squares (Fig. 3). The robust computation of the homography provides 55 matches, 11 of them being wrong as segments but good as infinite lines. Among the 105 final matches, there are less than 3% of wrong matches which correspond to contiguous cars. This means 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.

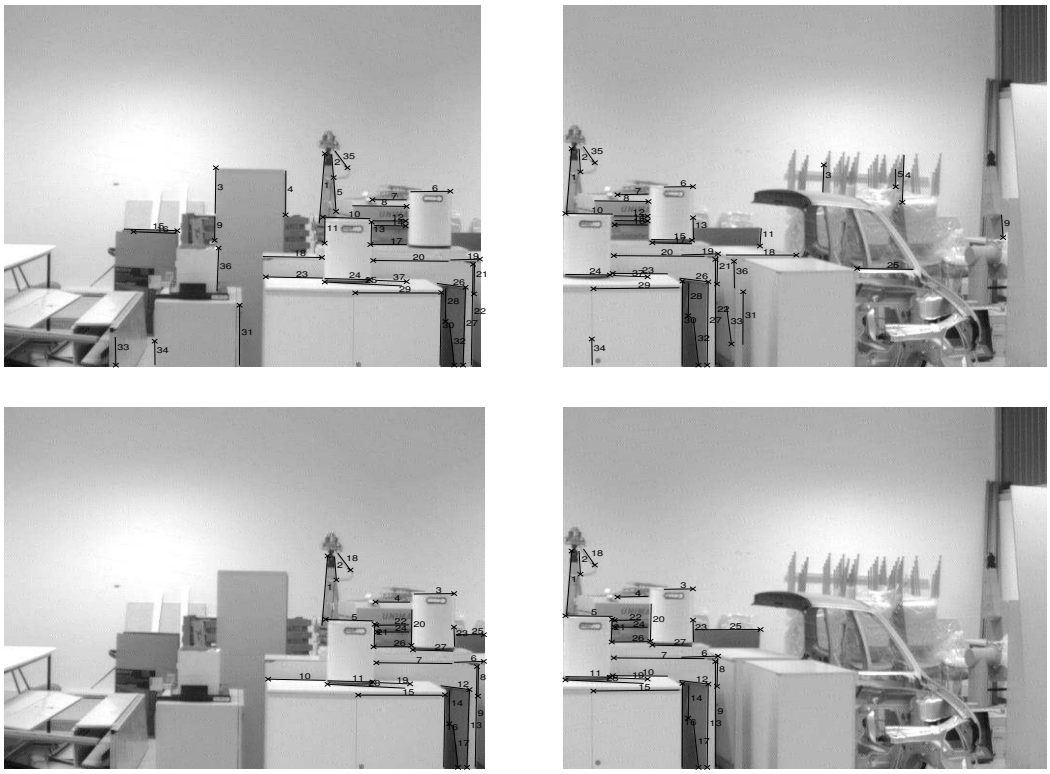

Figure 2: Images with 18 degrees of robot rotation. Images of first row show the matches after the basic matching (37 matches, 14 wrong). Second row shows the matches at the final stage of the process, only one been wrong match as segment although it is good as line (number 10).

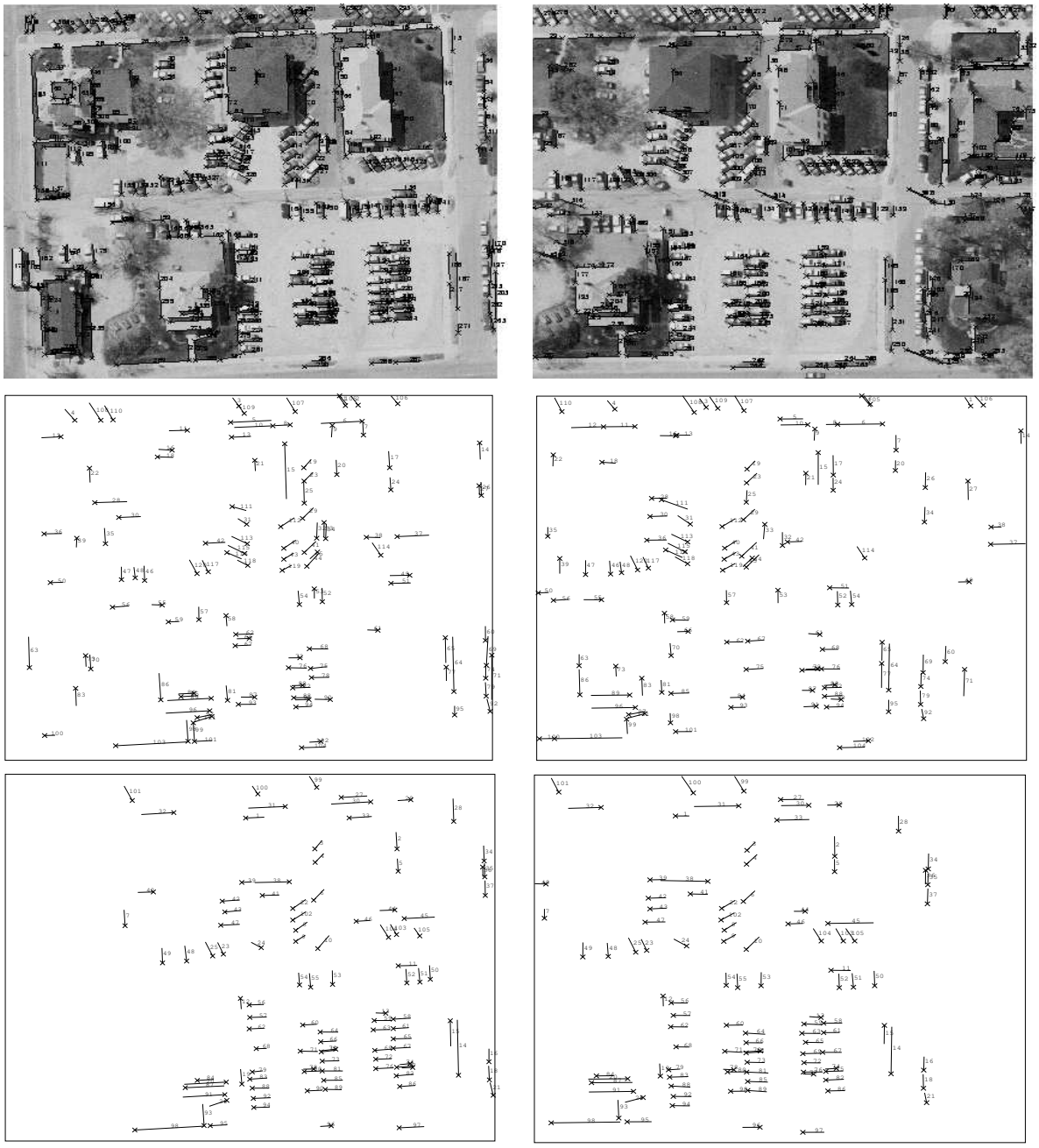

Figure 3: Two aerial images with large baseline. In the first row, the lines extracted are superimposed to the images (approximately 300 lines/image). The second row shows the basic matches (121 matches, 64 being wrong). Third row shows the matches at the final stage (105 matches, 3 being wrong that are corresponding to contiguous cars).

Figure 4: Two examples of image pairs with automatically matched vertical lines.

	Basic.Matches	Good	Final.Matches	Good
House	148	75%	114.6 (8.91)	99% (1%)
College	196	82%	156.7 (11.09)	96% (1%)

Table 2: Number of matches of scenes in Figs. 5 and 6. Only the lines on the two main planes are considered. We have repeated the experiments fifty times using the same basic matches, showing as final matches the mean values and, in brackets, the standard deviation.

We also show two examples of matched vertical lines with an homography of reduced dimension (Fig. 4). In both cases the percentage of good final matches is close to 100%. Using vertical lines, we have developed a real time implementation for robotic applications in man made environments [10].

Other experiments have been carried out to match lines which are in several planes, segmenting simultaneously the planes (Figs. 5, 6). In Table 2 we show the number of matches and the ratio of good matches for 50 repetitions of the robot computation. In this case, once an homography has been computed, the robust homography computation and the growing of matches process has been iteratively repeated twice. As it can be seen, the number and quality of the final matches is very good. The homology filter just commented in section 5 has been used to detect situations where a second plane is not obtained and therefore a sole homography can be computed, or when the homographies do not give a right homology due to noise or bad extraction. It can be seen that some lines of the second plane have been selected by the first homography (Fig. 6). The reason is that these lines are nearly coincident with epipolar lines and they can be selected by any homography being compatible with the fundamental matrix.

The method also work with rotated images. In this case σ_θ must be changed according to the image disparity. We show an example of a rotated image pair of short baseline where a sole homography gives very good results (Fig. 7).

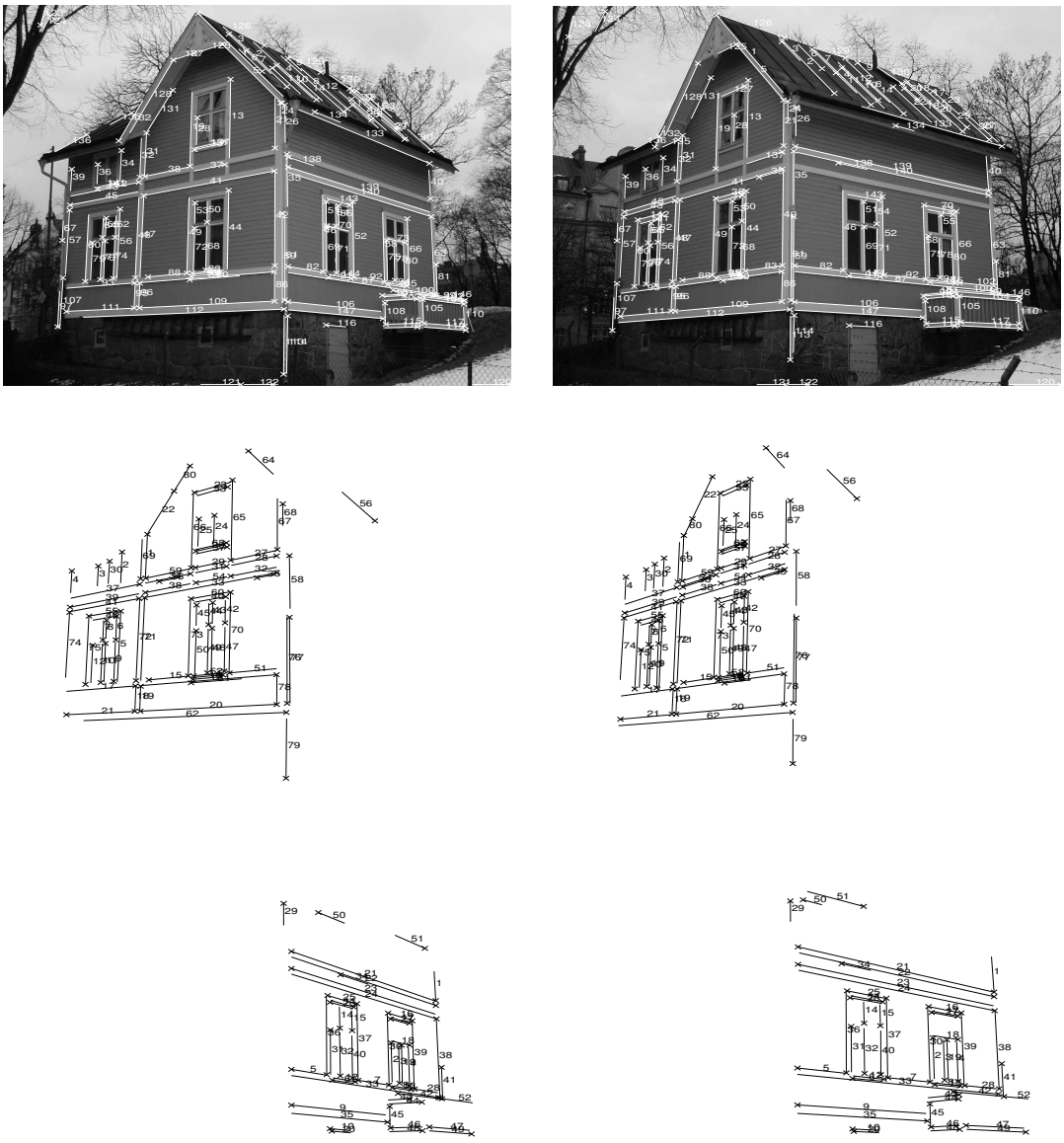

Figure 5: Images of the house. First row: basic matches. Second row: final matches corresponding to the first homography. Third row: final matches corresponding to second homography (Original images from KTH, Stockholm).

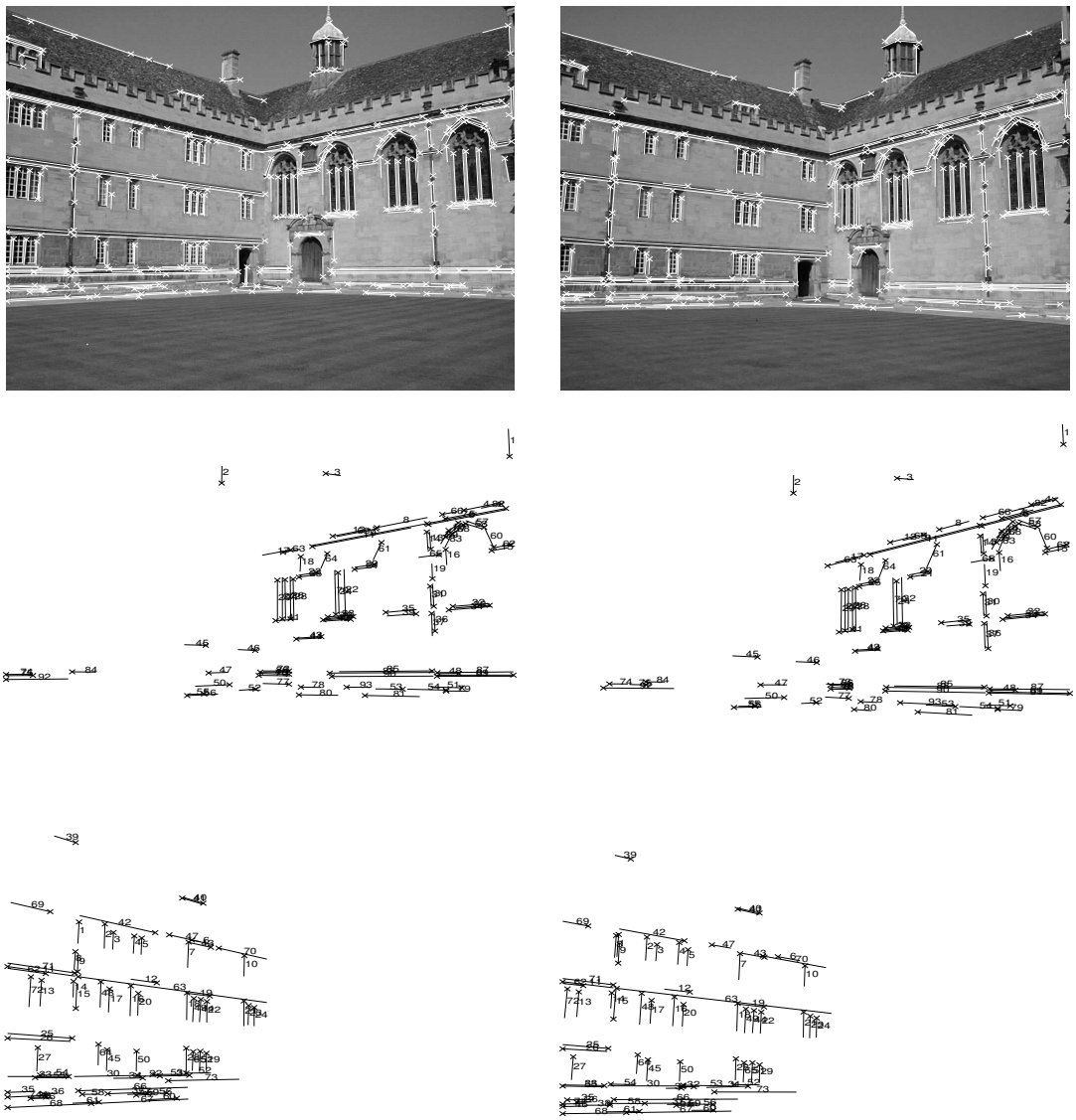

Figure 6: Images of the college. First row: basic matches. Second row: final matches corresponding to first homography. Third row: final matches corresponding to second homography (Original images from VGG, Oxford).

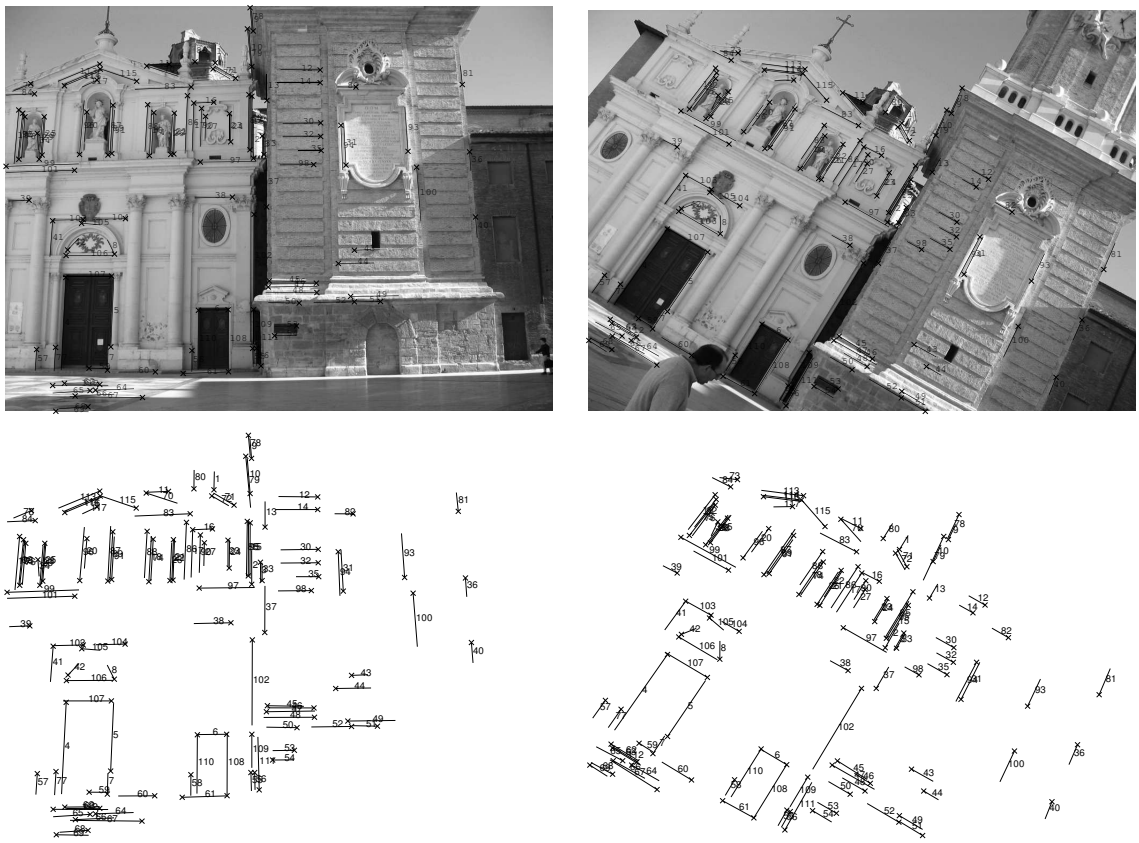

Figure 7: Pair of rotated short baseline images and final line correspondences.

8 Conclusions

We have presented and tested a method to automatically obtain matches of lines simultaneously to the robust computation of homographies in image pairs. The robust computation works especially well to eliminate outliers which may appear when the 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 automatic with only some previous tuning of parameters related to the expected global image disparity. As is seen in the experiments, the proposed algorithm works with different types of scenes. With planar scenes or in situations where disparity is mainly due to rotation, the method gives good results with only one homography. However, it is also possible to compute several homographies when several planes are in the scene, which are iteratively segmented. In this case, the fundamental matrix can be computed from lines through the homographies, obtaining a constraint for the whole scene. Currently we are working with other image invariants to improve the first stage of the process.

Acknowledgments

This work was supported by project DPI2003-07986.

References

[1] C. Schmid and A. Zisserman, "Automatic Line Maching across Views," In *IEEE Conference on CVPR*, pp. 666–671 (1997).

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision 60(2), 91–110 (2004).

[3] H. Bay, V. Ferrari, and L. V. Gool, "Wide-Baseline Stereo Matching with Line Segments," In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 329–336 (2005).

[4] P. Pritchett and A. Zisserman, "Wide Baseline Stereo Matching," In *IEEE Conference on Computer Vision*, pp. 754–760 (1998).

[5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University Press, Cambridge, 2000).

[6] Z. Zhang, "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] P. Torr and D. Murray, "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix," International Journal of Computer Vision 24, 271–300 (1997).

[8] L. Quan and T. Kanade, "Affine Structure from Line Correspondences with Uncalibrated Affine Cameras," IEEE Trans. on Pattern Analysis and Machine Intelligence 19(8), 834–845 (1997).

[9] R. Hartley, "In Defense of the Eight-Point Algorithm," IEEE Trans. on Pattern Analysis and Machine Intelligence 19(6), 580–593 (1997).

[10] J. Guerrero, R. Martinez-Cantin, and C. Sagüés, "Visual map-less navigation based on homographies," Journal of Robotic Systems 22(10), 569–581 (2005).

[11] A. Habib and D. Kelley, "Automatic relative orientation of large scale imagery over urban areas using Modified Iterated Hough Transform," Journal of Photogrammetry and Remote Sensing 56, 29–41 (2001).

[12] C. Sagüés, A. Murillo, F. Escudero, and J. Guerrero, "From Lines to Epipoles Through Planes in Two Views," Pattern Recognition (2006), to Appear.

[13] J. Guerrero and C. Sagüés, "Robust line matching and estimate of homographies simultaneously," In *IbPRIA, Pattern Recognition and Image Analysis, LNCS 2652*, pp. 297–307 (2003).

[14] J. Burns, A. Hanson, and E. Riseman, "Extracting Straight Lines," IEEE Trans. on Pattern Analysis and Machine Intelligence 8(4), 425–455 (1986).

[15] R. Deriche and O. Faugeras, "Tracking Line Segments," In *First European Conference on Computer Vision*, pp. 259–268 (Antibes, France, 1990).

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

[17] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Comm. of the ACM 24(2), 381–395 (1981).

[18] L. Zelnik-Manor and M. Irani, "Multiview Constraints on Homographies," IEEE Transactions on Pattern Analysis and Machine Intelligence 24(2), 214–223 (2002).