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

Internal Report: 2005-V03

Automatic Matching and Motion Estimation from Two Views of a Multiplane Scene[1]

G. López-Nicolás, C. Sagüés, J.J. Guerrero

If you want to cite this report, please use the following reference instead:
Automatic Matching and Motion Estimation from Two Views of a Multiplane Scene, G. López-Nicolás, C. Sagüés, J.J. Guerrero, *Pattern Recognition and Image Analysis*, LNCS 3522, pages 69-76, 2005.

[1] This work was supported by project DPI2003-07986.

Automatic Matching and Motion Estimation from Two Views of a Multiplane Scene

G. López-Nicolás, C. Sagüés, and J.J. Guerrero

Dpto. de Informática e Ingeniería de Sistemas
Instituto de Investigación en Ingeniería de Aragón, Univ. de Zaragoza
Edificio Ada Byron, C/ María de Luna 1, E-50018 Zaragoza, Spain
gonlopez, csagues, jguerrer@unizar.es *

Abstract. This paper addresses the computation of motion between two views when 3D structure is unknown but planar surfaces can be assumed. We use points which are automatically matched in two steps. The first one is based on image parameters and the second one is based on the geometric constraint introduced by computed homographies. When two or more planes are observed, corresponding homographies can be computed and they can be used to obtain the fundamental matrix, which gives constraints for the whole scene. The computation of the camera motion can be carried out from a homography or from the fundamental matrix. Experimental results prove this approach to be robust and functional for real applications in man made environments.

Keywords: Matching points, multiplane scenes, homographies, fundamental matrix, motion estimation.

1 Introduction

The fundamental matrix encapsulates the geometric information which relates two different views regardless of the observed scene. The non metric basis of this matrix makes possible to use uncalibrated cameras. It has been usually computed through points [1] although lines can also be used when two or more planes are available [2]. Obviously points can also be used to compute homographies and, if two or more homographies are available, the fundamental matrix can be computed from them [3], [4].

In all the cases the matching problem is crucial to make the process work automatically. The matching of features based on image parameters may give non matched or wrong matched features. Projective transformations allow image dependent measures, as cross-correlation, to be a viewpoint invariant, which make possible to afford wide baseline matching [5]. So, the constraint imposed by fundamental matrix or homographies must be used for matching points.

Scenes with several planes are usual in man made environments, and the model to work with multiple views of them is well known. Points or lines in one

* This work was supported by project DPI2003-07986.

image of the world plane are mapped to points or lines in the other image by a plane to plane homography [6]. We robustly match points between two images using the projective transformations corresponding to the existing scene planes. The robust matching of points and the computation of the corresponding homography is iteratively carried out until we have no more available planes. If two planes have been computed at least, the fundamental matrix can be computed, which gives general constraint for the whole scene. It has been reported that the multi-plane algorithm is not as stable as the general method [3], but when less than three planes are observed, which is quite usual in man made environments, the multi-plane algorithm gives better results than the general method.

Camera motion between two views can be obtained from the computed homography or from the fundamental matrix. Both methods are exposed in this paper. Normally the computation of motion has been directly considered from the fundamental matrix, which is a more general model. However, the fundamental matrix is ill conditioned with short baseline or when all the points lie on a plane, which may easily happen in man made environments [6]. In these cases the fundamental matrix is an inappropriate model to compute camera motion. Using homographies, we can check the homology conditioning to determine if the fundamental matrix may be computed. Therefore we can choose the appropriate motion algorithm from either the fundamental matrix or the homography.

2 Robust Matching

Automatic matching continues to be an unsolved problem in general situations. The aim is to determine correspondences between points in two images without knowledge about motion or scene structure.

In this work the points of interest are extracted with the Harris corner extractor [7]. To obtain a homogeneous distribution of points all over the image, it is divided in a grid and we establish a maximum number of points per cell to be extracted. Additionally we establish a threshold of minimum contrast just to give only good points.

Later, we consider the matching in two steps, the first step is based on image correlation on a search window around the candidate points. This is actually the most weak step of our implementation because, as known, correlation is not invariant to rotations. As some mismatches appear here, we introduce in the second step, our "friendship" algorithm. It is similar to the previously proposed relaxation process [8]. The idea is to allow only the matches whose neighboring points move similarly. Those that do not behave as the neighbors are eliminated.

These points can be represented in the projective plane with homogeneous coordinates as $\mathbf{p} = (x, y, 1)^T$. A projective transformation $\mathbf{H}_{21}$ exists from matched points belonging to a plane in such a way that $\mathbf{p}_2 = \mathbf{H}_{21}\mathbf{p}_1$.

From the previous relation each couple of corresponding points gives two homogeneous equations to compute the projective transformation, which can be determined up to a non-zero scale factor. To compute the homography, we have chosen the RANSAC method [9], which is a robust method to consider the existence of outliers. It makes a search in the space of solutions obtained from subsets of four matches. Each subset provides a 8×9 system of equations whose solution is obtained from singular value decomposition.

From here on, we introduce the geometrical constraint introduced by the estimated homography to get a bigger set of matches. Thus, final matches are composed by two sets. The first one is obtained from the matches selected after the robust computation of the homography. The second one is obtained making a rematching of not matched points based on the computed homography.

3 From Homographies to Fundamental Matrix

Fundamental matrix has been stated as a crucial tool when using uncalibrated images. As known, it is a 3×3 matrix of rank 2 which encapsulates the epipolar geometry. It only depends on internal parameters of the camera and the relative motion.

Let us suppose the images are obtained with the same camera whose projection matrixes in a common reference system are $\mathbf{P}_1 = \mathbf{K}[\mathbf{I}|\mathbf{0}], \mathbf{P}_2 = \mathbf{K}[\mathbf{R}|\mathbf{t}]$; being $\mathbf{R}$ the camera rotation, $\mathbf{t}$ the translation and $\mathbf{K}$ the internal calibration matrix. Then, the fundamental matrix can be expressed as $\mathbf{F}_{21} = \mathbf{K}^{-T}\ ([\mathbf{t}]_\times \mathbf{R})\ \mathbf{K}^{-1}$. Normally, it has been computed from corresponding points [1], [10], using the epipolar constraint, which can be expressed as $\mathbf{x}_2^T\, \mathbf{F}_{21}\, \mathbf{x}_1 = 0$. However, the fundamental matrix is unstable when points lie in a plane [10]. In [3] is shown that the multiplane method behaves better than the general method when less than three planes are available. This constrained structure is usually observed in man made environments.

In the case of multiplane scenes some alternatives can be used to compute the fundamental matrix. If at least two homographies $(\mathbf{H}_{21}^{\pi_1}, \mathbf{H}_{21}^{\pi_2})$ corresponding to two planes (π_1, π_2) can be computed between both images, the homology on the second image $\mathbf{H}_2 = \mathbf{H}_{21}^{\pi_1} \cdot (\mathbf{H}_{21}^{\pi_2})^{-1}$, which is a mapping from one image onto itself, can be computed. Under this mapping the epipole is a fixed point $\mathbf{e}_2 = \mathbf{H}_2\, \mathbf{e}_2$, so it may be determined from the eigenvector of $\mathbf{H}_2$ corresponding to non unary eigenvalue [6]. Therefore, the fundamental matrix can be computed using $\mathbf{H}_{21}^{\pi_1}$ or $\mathbf{H}_{21}^{\pi_2}$ as,

$$\mathbf{F}_{21} = [\mathbf{e}_2]_\times\, \mathbf{H}_{21}^{\pi_i} \ , \tag{1}$$

being $[\mathbf{e}_2]_\times$ the skew matrix corresponding to $\mathbf{e}_2$ vector.

On the other hand, the fundamental matrix can also be computed from both homographies through a system of twelve linear equations extracted from the following relation [3],

$$\mathbf{H}_{21}^{\pi_i\,T}\, \mathbf{F}_{21} + \mathbf{F}_{21}{}^T\, \mathbf{H}_{21}^{\pi_i} = 0 \ . \tag{2}$$

As we propose to compute fundamental matrix from homographies, a check on the homology conditioning may help to determine if the fundamental matrix may or may not be computed. Similarly the homology on the first image can be computed as $\mathbf{H}_1 = (\mathbf{H}_{21}^{\pi_1})^{-1} \cdot \mathbf{H}_{21}^{\pi_2}$ and taking into account that for a plane $\mathbf{H}_{21} = \mathbf{K}\,(\mathbf{R} - \frac{\mathbf{t}\,\mathbf{n}_\pi{}^T}{d_\pi})\,\mathbf{K^{-1}}$, it turns out that the eigenvalues of the $\mathbf{H}_1$ homology are $(1, 1, 1 + \mathbf{v}^T\mathbf{p})$ being $\mathbf{v} = \mathbf{K}\mathbf{R}^{-1}\mathbf{t}/(1 - \frac{\mathbf{n}_{\pi_1}^T}{d_{\pi_1}}\mathbf{R}^{-1}\mathbf{t})$ a view dependent vector, and $\mathbf{p} = (\frac{\mathbf{n}_{\pi_1}^T}{d_{\pi_1}} - \frac{\mathbf{n}_{\pi_2}^T}{d_{\pi_2}})\mathbf{K^{-1}}$ a plane dependent vector, being $\mathbf{n}_{\pi_1}$, $\mathbf{n}_{\pi_2}$ the normals and d_{π_1}, d_{π_2} the distances of the planes [11].

So, the homology has two equal eigenvalues. The third one is related to the motion and the structure of the scene. These eigenvalues are used to test when two different planes have been computed, and then the epipole and the intersection of the planes can be also computed. The epipole is the eigenvector corresponding to the non-unary eigenvalue and the other two eigenvectors define the intersection line of the planes [6]. In case of small baseline or if there is only one plane in the scene, epipolar geometry is not defined and only one homography can be computed, so possible homology $\mathbf{H}_1$ will be close to identity, up to scale.

In practice a filter is proposed 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 is rejected. On the other hand, if the three eigenvalues are similar we check if the homology is close to identity to avoid the case where two similar homographies are computed.

4 Camera Motion from Two Views

Complete motion (rotation and translation up to a scale factor) can be computed from homography or from the fundamental matrix if camera is calibrated. As we have seen before, the homography $\mathbf{H}_{21}$ can be related to motion in such a way that $\mathbf{H}_{21} = \mathbf{K}\,(\mathbf{R} - \frac{\mathbf{t}\,\mathbf{n}^T}{d})\,\mathbf{K^{-1}}$, being $\mathbf{n}$ the normal to the scene plane and d its depth. From here, two solutions (up to a scale factor for $\mathbf{t}$) can be obtained [12]. The main steps of this algorithm is summarized in Algorithm 1.

Algorithm 1 Motion algorithm from homography

---

1. Compute a calibrated homography $\mathbf{H^c_{21}} = \mathbf{K^{-1}}\,\mathbf{H}_{21}\,\mathbf{K}$
2. Compute the singular value decomposition of matrix $\mathbf{H^c_{21}}$, in such a way that $\mathbf{H^c_{21}} = \mathbf{U}\,diag(\lambda_1, \lambda_2, \lambda_3)\,\mathbf{V}^T$ with $\lambda_2 = 1$
3. Let be $\mathbf{S}^T\,\mathbf{S} = diag(\lambda_1,\, \lambda_2,\, \lambda_3)$, and $\alpha = \sqrt{\frac{\lambda_3-\lambda_2}{\lambda_3-\lambda_1}}$, $\beta = \sqrt{\frac{\lambda_2-\lambda_1}{\lambda_3-\lambda_1}}$
4. Writing $\mathbf{V} = [\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3]$, compute $\mathbf{v}_v = \alpha\,\mathbf{v}_1 + \beta\,\mathbf{v}_3$
5. Compute rotation matrix $\mathbf{R} = [\mathbf{H^c_{21}}\,\mathbf{v}_v,\ \mathbf{H^c_{21}}\,\mathbf{v}_2,\ \mathbf{H^c_{21}}\,\mathbf{v}_v \times \mathbf{H^c_{21}}\,\mathbf{v}_2][\mathbf{v}_v, \mathbf{v}_2, \mathbf{v}_v \times \mathbf{v}_2]^T$
6. Compute translation up to a scale factor as $\mathbf{t} = \mathbf{H^c_{21}}\,\mathbf{n} - \mathbf{R}\,\mathbf{n}$ being $\mathbf{n} = \mathbf{v}_v \times \mathbf{v}_2$
7. The second solution for $\mathbf{R}$ and $\mathbf{t}$ can be obtained by making $\beta = -\beta$
8. If $\lambda_3 = \lambda_2$, there is a sole solution being the camera translation perpendicular to the plane ($\mathbf{t} \parallel \mathbf{R}\,\mathbf{n}$) and coming nearer the plane. If $\lambda_1 = \lambda_2$ there is also a sole solution, but now the camera gets away from the plane. Finally, if $\lambda_1 = \lambda_2 = \lambda_3$ report the sole solution $\mathbf{t} = 0$, and $\mathbf{R} = \mathbf{H^c_{21}}$

Algorithm 2 Motion algorithm from fundamental matrix

1. Compute the essential matrix $\mathbf{E} = \mathbf{K}^T\,\mathbf{F}\,\mathbf{K}$
2. Compute the singular value decomposition of matrix $\mathbf{E}$, in such a way that $\mathbf{E} = \mathbf{U}\,diag(1,1,0)\,\mathbf{V}^T$
3. The camera translation, up to a scale factor is $\mathbf{t} = \mathbf{U}\,(0,0,1)^T$
4. The two solutions for the rotation matrix are $\mathbf{R} = \mathbf{U}\,\mathbf{W}\,\mathbf{V}^T$ and $\mathbf{R} = \mathbf{U}\,\mathbf{W}^T\,\mathbf{V}^T$, being $\mathbf{W} = \left[(0,1,0)^T, (-1,0,0)^T, (0,0,1)^T\right]$

Camera motion can also be computed from the fundamental matrix. As in previous case, the algorithm provides two solutions up to a scale factor for translation. Given the calibration matrix, the motion can be deduced from $\mathbf{F}$ as summarized in Algorithm 2 [6].

In case of pure rotation or if there exists only one plane in the scene, the epipolar geometry is not defined. Then, only the alternative of motion from homography will be correct.

5 Experimental Results

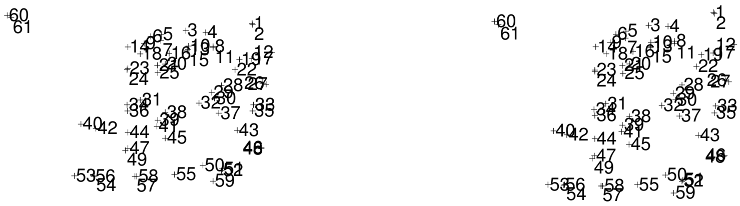
Many experiments have been carried out with synthetic and real images. The homology filter just commented has been used to determine when a second plane has been obtained. Several criteria can be used to measure the accuracy of the computed motion. With synthetic images, where motion is known, we measure the rotation error. We also measure the first order geometric error computed as the Sampson distance [6] for a set of corresponding points manually extracted and matched.

With real images the matches are automatically obtained for two planes in scene (Fig. 1). The points extracted are 479 from the first image and 475 from the second. The number of basic matches obtained is 147 with 86.4% of good matches. Once a homography has been computed, the robust homography computation and the growing matches process has been iteratively repeated twice. The experiment has been repeated 50 times using the same basic matches, and the mean of final matches obtained is 131.8 matches ($\sigma = 10.5$) with 96.9% of good matches ($\sigma = 1.2\%$). As it can be seen the number and quality of final matches are quite good.

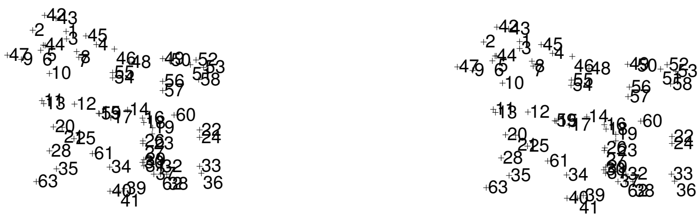
As we have seen, one of the results of the homology is the intersection line of the planes. We have proposed to use a filter based on the homology eigenvalues to avoid situations where a sole homography can be computed or where the homographies do not give a right homology due to noise or bad extraction. In these cases the epipole, the fundamental matrix or the intersection line would be badly computed. In Fig. 2 we can see the intersection lines of the planes for 100 executions with and without the homology filter. As it can be seen the quality of the results improves significantly with the proposed filter.

With respect to the fundamental matrix computation, we show (Table 1) the mean of the Sampson distance for 20 points manually extracted and matched.

(a) (b)



(c) (d)



(e) (f)

Fig. 1. Images of the college to compute homographies. Extracted points (a), (b). Matches corresponding to the first homography (c), (d) and to the second (e), (f). (Original images from VGG, Oxford)

(a) (b)

Fig. 2. Intersection of the planes through the eigenvalues of the homology. The lines corresponding to 100 executions are represented without filter (a), and with homology filter (b)

Table 1. Sampson distance for 20 points manually matched (belonging to each plane for homographies and distributed around the scene for fundamental matrixes). We show in 100 executions the median and the mean with and without filter. These results are shown for the homographies (H1, H2) and for the fundamental matrixes: eH1 and eH2 using (1) with $\mathbf{H}_{21}^{\pi_1}$ and $\mathbf{H}_{21}^{\pi_2}$ respectively, and FH using (2)

		Synthetic (pixels)					Oxford college (pixels)				
		H1	H2	eH1	eH2	FH	H1	H2	eH1	eH2	FH
Without filter	median	0.581	0.586	0.891	0.789	0.932	0.707	0.683	1.004	1.286	1.906
	mean	0.577	0.586	1.619	1.458	1.634	0.709	0.698	4.998	5.187	12.61
With filter	median	0.581	0.584	0.740	0.725	0.805	0.687	0.666	0.566	0.796	1.045
	mean	0.578	0.587	0.926	0.767	0.883	0.697	0.694	0.642	0.789	1.099

We consider the images of the college and two synthetic images. The synthetic scene consists of random points, with white noise of $\sigma = 0.3$ pixels, distributed in three perpendicular planes. The experiment has been repeated 100 times and we show mean and median values. The Sampson distance is similar for the three presented ways of computing the fundamental matrix, although it is a bit worse using (1). Probably this is because if one homography is less accurate than the other, (2) collects this inaccuracy, currently we are studying the implications of these differences.

Table 2. Mean of rotation error (Synthetic) and rotation angle (College) computing motion through homographies H1 or H2 with algorithm 1, and through fundamental matrixes, eH1 and eH2 using (1) and FH using (2), with algorithm 2

	Synthetic: rotation error (deg)					Oxford college: rotation (deg)				
	H1	H2	eH1	eH2	FH	H1	H2	eH1	eH2	FH
Without filter	0.958	0.454	0.524	0.545	0.562	9.240	10.64	7.777	7.662	8.096
With filter	0.456	0.365	0.225	0.226	0.214	9.691	10.97	9.118	9.115	9.478

Finally, results of the computation of camera motion using homographies and fundamental matrix are exposed. We have executed these algorithms 100 times. Table 2 shows the mean of the rotation (Oxford college) and the rotation error (synthetic data) obtained through homographies (Algorithm 1) and fundamental matrixes (Algorithm 2). Fundamental matrix is computed in different ways using equations (1) and (2). The results are exposed with and without the homology filter and they show the goodness of the proposed filter.

6 Conclusions

We have presented the matching of points, the computation of the intersection of the planes and the computation of camera motion from two views. This is carried out through homographies corresponding to planes, which are quite usual in man made environments. The robust computation of matches based on homographies works especially well to automatically eliminate outliers which may appear when there is no information of scene structure or camera motion. The fundamental matrix and the intersection line of the planes is properly obtained if the images correspond to motion and scenes which are geometrically well conditioned. If it does not happen a homography may be given as a result of the algorithm and motion can be obtained from this homography.

The main achievement of this work is that all the process is made automatically and works in a robust way. Besides this, the joint use of homographies and fundamental matrix allows the properly selection of the model to determine camera motion in real applications. The proposed approach is a good solution in man made environments, where usually at least one plane is available.

References

1. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial Intelligence **78** (1995) 87–119
2. Pellejero, O., Sagüés, C., Guerrero, J.: Automatic computation of fundamental matrix from matched lines. In: Current Topics in Artifical Intelligence, LNCS-LNAI 3040. (2004) 197–206
3. Luong, Q.T., Faugeras, O.: Determining the fundamental matrix with planes: Unstability and new algorithms. In: Proc. Conference on Computer Vision and Pattern Recognition, New-York (1993) 489–494
4. Rother, C., Carlsson, S.: Linear multi view reconstruction and camera recovery using a reference plane. International Journal of Computer Vision **49** (2002)
5. Pritchett, P., Zisserman, A.: Wide baseline stereo matching. In: IEEE Conference on Computer Vision. (1998) 754–760
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
7. Harris, C., Stephens, M.: A combined corner and edge detector. The Plessey Company (1988)
8. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Rapport de recherche RR-2273, I.N.R.I.A., Sophia-Antipolis, France (1994)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM **24** (1981) 381–395
10. Luong, Q., Faugeras, O.: The fundamental matrix: Theory, algorithms, and stability analysis. Int. Journal of Computer Vision **17** (1996) 43–76
11. Zelnik-Manor, L., Irani, M.: Multiview constraints on homographies. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 214–223
12. Weng, J., Huang, T., Ahuja, N.: Motion and Structure from Image Sequences. Springer-Verlag, Berlin-Heidelberg (1993)