Planar facets segmentation using a multiresolution dense disparity field estimation

Lionel Oisel, Luce Morin, Étienne Mémin and Claude Labit

IRISA/INRIA
Campus Universitaire de Beaulieu, 35042 Rennes Cedex, FRANCE
e-mail:{loisel,lmorin,memin,labit}@irisa.fr

Abstract

In the present paper we propose a new algorithm for planar facets segmentation of sequences of uncalibrated images in order to recover 3D models of complex scenes. This is performed using a two steps algorithm. First a robust and regularized dense disparity field is computed under the epipolar geometry constraint. The resulting field is segmented according to homographic models using an iterative Delaunay triangulation.

1 Introduction

The general issue of our work is the recovering of 3D information from sets or sequences of uncalibrated images, in the context of a static scene, or equivalently images taken simultaneously from different view-points. This issue is of interest in many applications, for instance generating intermediate or extrapolated views for object manipulation in multimedia applications.

Many recent works try to deal with this subject. All of them allows a 3D reconstruction that can be projective or Euclidean. Two different approaches are generally proposed using different 2D information:

- dense approach: a dense disparity field is computed. The resulting field is triangulated to obtain a three dimensional cluster of points [6][8]. The main problem of these approaches is to obtain a smooth and accurate dense map so that the 3D reconstruction would contain unpleasant outliers.

- scattered approach: 3D information composed of polygonal objects is described by a limited set of 3D points[2][4] which characterize the vertices of each 3D plan. It is usually necessary to have a manual interaction to obtain these points. This method is particularly efficient for architectural reconstruction where the number of characteristic points is limited. The snag with such approaches is that the number of vertices dramatically increases in the case of complex scenes.

We also want to allow an interactive real time visualization for the recovered information. This is obtained by providing a description of the 3D scene as a triangular mesh, which can be fairly displayed by most visualization dedicated systems. Our method thus belongs to the second type (scattered approach). As we want to process complex scenes, the modelization must be computed automatically.

The key point of our method is to segment the images into regions of planar points which is equivalent to motion segmentation according to an homographic model. As the homographic model associated to planar facets is non linear, a region-based method is hardly feasible. We decided to split the algorithm in two different steps. The first one will give us a dense depth map associated to a discontinuity map. This dense information is used to initialize the second step: homographic model estimation and segmentation.

The first step of this scheme consists in estimating a well constrained disparity field in order to facilitate the segmentation step. To do so, classical techniques based on correlation are not suitable. An algorithm derived from the optical flow estimation problem is used, providing a robust and regularized estimated field [7][9] which respects the so-called epipolar geometry. This insures that the estimated field is geometrically consistent with a perspective projection model and with the hypothesis of a fixed scene. This added constraint also reduces the computational cost.

The second step uses the resulting epipolar constrained disparity field to perform triangulation of the two images. Each corresponding triangle pair is associated to an homographic model.

2 Dense disparity field estimation

2.1 General principle

Let $I_i(s)$ be the intensity in the ith image, where $s = (x, y)$ denotes spatial position. Under the assumption of constant intensity along motion trajectories, the DFD (Displaced Frame Difference) is equal to zero:

$$DFD(s, d_s) = I_1(s) - I_2(s + d_s) \ = \ 0 \ , \qquad (1)$$

where $d_s = (d_x, d_y)$ stands for the image displacement from position 1 to position 2 of a material point along x and y axis. In the general case, this is a 2D problem: for each pixel, d_x and d_y have to be found. Using the *epipolar constraint* the number of unknowns can be reduced to solve a 1D problem. This constraint is entirely defined by a 8 parameters homogeneous matrix called the *fundamental matrix*, which can be computed from uncalibrated pairs of stereoscopic images. This requires at least 8 point matches which can be automatically extracted [11]. For any point in the first image, the fundamental matrix provides the line of potential correspondent points in the second image (*epipolar line*):

$$\tilde{l} = F.\tilde{s}$$

where F is the fundamental matrix, $\tilde{}$ denotes homogeneous coordinates, l the epipolar line associated to a point s.

The fundamental matrix F is recovered from automatically extracted and matched points on the original pair of images. The estimation method is based on a virtual parallax algorithm [5][3].

Using the epipolar constraint, we can decompose the displacement vector d_s into normal and tangent components with respect to the epipolar line (see Figure 1) leading to the following equation: $DFD(s, d_s) = I_1(s) - I_2(s + \vec{N}_s + \lambda_s \vec{V}_s) = 0$. The normal component $\vec{N}_s$ and the unit vector on the epipolar line $\vec{V}_s$ can be computed for any position s using the fundamental matrix. Enforcing the epipolar constraint for every point, the original problem of a 2D displacement field estimation is now reduced to the 1D problem of estimating the disparity λ_s along the epipolar line.

2.2 Multiresolution scheme

The fundamental matrix can be reliably estimated only for large displacements between the 2 camera view-points. On the other hand, Equation (2) is solved by linearization with respect to λ_s, thus assuming small displacement. To overcome this incompatibility, the estimation is embeded in a coarse-to-fine multiresolution scheme. At a given level k, the disparity λ_s^k is decomposed into a previously estimated disparity λ_s^{k-1} from coarser level $k-1$ and a refinement $d\lambda_s^k$ to be estimated, leading the following equation:

$$I_1^k(s) - I_2^k(s + \vec{N}_s^k + \lambda_s^{k-1}\vec{V}_s^k + d\lambda_s^k \vec{V}_s^k) = 0 \qquad (2)$$

to be solved with respect to $d\lambda_s^k$.

Pyramids of images I_1^k, I_2^k, and pyramids of tangent and normal vectors $\vec{N}_s^k$, $\vec{V}_s^k$ are built previously to the estimation.

A fundamental matrix F^k is computed for each level from the initial matrix F and a change of coordinates. F^k allows to compute $\vec{N}_s^k$ and $\vec{V}_s^k$ for each position s. To ensure that the epipolar constraint is respected, the displacement vector estimated at level $k-1$ is projected onto the epipolar line (see Fig. 2).

2.3 Regularized estimation method

We now consider a given level k. Upperscript k will be omitted for clarity. Equation (2) is linearized with respect to $d\lambda_s$ around $s + \vec{N}_s + \lambda_s \vec{V}_s$. $d\lambda_s$ is considered as a realization of a Markov random field. The best disparity field according to the maximum a posteriori (MAP) Bayesian criterion is given by:

$$\widehat{d\lambda} = \arg\min_{d\lambda \in \mathbb{R}} \ H(d\lambda) \ = \ \arg\min_{d\lambda \in \mathbb{R}} \ (H_1(d\lambda) + \alpha[H_2(d\lambda)]) \qquad (3)$$

where α is an arbitrary fixed real. H_1 is the observation function:

$$H_1(d\lambda) = \sum_s \left[d\lambda_s \vec{V}_s \nabla \tilde{I}_2(s) + \tilde{I}_2(s) - I_1(s) \right]^2$$

$$\text{with} \ \ \tilde{I}_2(s) = I_2(s + \vec{N}_s + \lambda_s \vec{V}_s)$$

where $\nabla \tilde{I}_2(s)$ is the spatial gradient on the displaced image $\tilde{I}_2(s)$. H_2 is the smoothness term which favors similar displacement vectors d_s and d_r for all pairs $< s, r >$ of neighbor positions:

$$H_2(d\lambda) = \sum_{<s,r>} \rho(d_s - d_r)$$

To cope with large deviations from the data model (resp. to authorize disparity depth discontinuities),

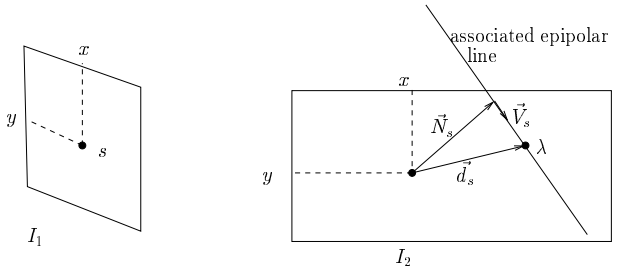

Figure 1: displacement vector decomposition

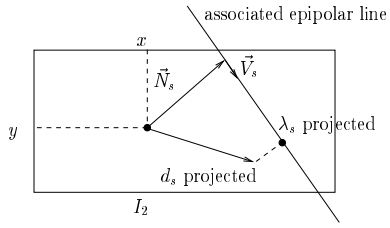

Figure 2: The vector projection according to epipolar geometry

H_1 (resp. H_2) includes a M–estimator ρ. Under certain conditions [1], H_1 and H_2 can be expressed as a quadratic function of $d\lambda_s$. The energy contribution of a point s to H_1 is weighted by a factor δ_s: the larger the contribution, the smaller the weight. Similarly, each pair of neighbors $< s, r >$ contributes to H_2 with a weight β_{sr} depending on their displacement vector difference $\|d_s - d_r\|$. The larger the difference, the smaller the weight.

At each given level, an iterative Gauss-Seidel scheme is used to solve the minimization problem. The minimization is conducted alternatively on the disparity field $d\lambda_s$ and on weight fields δ and β. Considering a fixed disparity field, weights are updated in closed form depending on the derivative of the used robust function. Now considering weights as being frozen, the disparity field is updated at each site taking the local optimum deduced from equation (3):

$$d\lambda_s = \frac{\alpha(-\lambda_s \sum_{<s,r>} \beta_{sr} + \vec{V}_s.\vec{\omega}_s) - \delta_s.\vec{V}_s.\nabla\tilde{I}_2(s).I_t(s)}{\delta_s.(\vec{V}_s.\nabla\tilde{I}_2(s))^2 + \alpha \sum_{<s,r>} \beta_{sr}} \quad (4)$$

where ω_s is the weighted average of neighboring disparity vectors. An initial disparity field is necessary: it is estimated on the original images using the set of matched points of interest and interpolation based on Delaunay triangulation. This field is projected on the top level of the pyramid to provide an initial disparity field for the low resolution level.

3 Segmentation

The segmentation step includes two cooperative operations: the triangulation and the computation of the homography matrices associated to each triangle.

3.1 Homography estimation

Given one triangle and the disparity field the goal is to compute the best homography model for all points in the triangle. The homography estimation is performed using a method proposed by Robert [10]. This method takes into account the epipolar geometry (expressed by the 3×3 matrix F) to efficiently compute the homography matrix from two or more corresponding pairs of points. Let $H(3 \times 3)$ be an homogeneous homography matrix and m_l and m_r be corresponding points on left image l and right image r. For H to be consistent with epipolar geometry, the homogeneous symmetric matrix $F^T.H + H^T F$ must be null, leading to 6 homogeneous equations with unknowns h_{ij}. Each corresponding pair of points (m_l, m_r) gives one scalar equation:

$$[m_r^i, Fm_l^i, Hm_l^i] = 0$$

where $[a, b, c]$ denotes the triple product. This leads to a system of the following form:

$$\begin{cases} F^T.H + H^T & = & 0 \\ [m_r^i, Fm_l^i, Hm_l^i] & = & 0 \\ \vdots & & \end{cases} \quad (5)$$

This can be rewritten as $Ah = 0$, where h contains the unknown coefficients of H. An estimate of h is computed using a SVD (singular value decomposition) of the matrix $A^t A$.

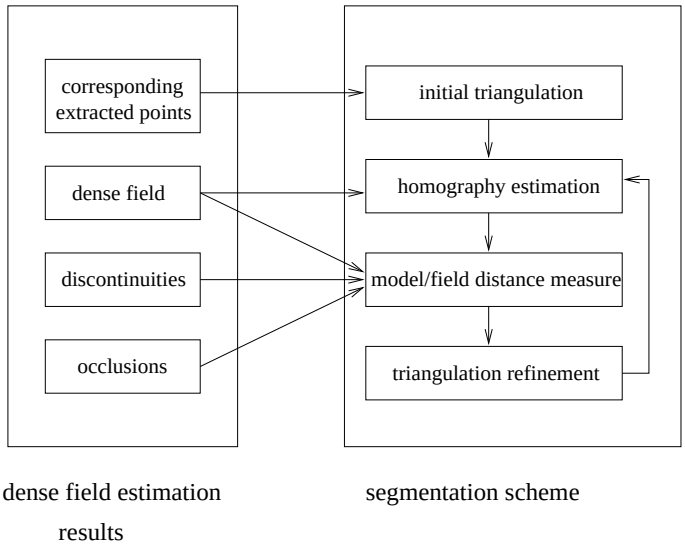

Figure 3: Segmentation scheme

3.2 Iterative triangulation

Figure 3 shows how the information resulting from the dense field estimation is used to compute the iterative triangulation.

An initial triangulation is performed from the set of corresponding points extracted for the fundamental matrix estimation. An iterative process is then used to refine this triangulation. For each triangle the associated homography matrix is computed using dense information and F.

A measure of the adequacy of H to the disparity field is then computed. The influence of each point s of the triangle is weighted by δ_s (occlusions areas do not influence the distance measurement). This measure also takes into account the discontinuity information: a triangle is cut taking into account the number of important discontinuity values it contains. For each new triangle the associated homography matrix is then computed.

The iterative triangulation refinement is performed until the distance measure is less than an arbitrary fixed threshold ϵ.

4 Results

The proposed method has been applied on different images including stereo pairs and images taken from a static scene viewed by a moving camera. Concerning the dense disparity estimation method, results have been compared to those given by a non constrained optical flow estimator. Due to the large displacements, the non constrained method (even using a multiresolution and multigrid approach) does not converge towards an acceptable solution.

Figure 5 presents two images of an indoor sequence taken by a commercial camera. Epipolar, initial interpolation and triangulation have been estimated from 122 automatically extracted matching points. On the discontinuity map, important discontinuity areas lie on the table and cabinet edges. The observation map characterizes important gradient areas and occultation zones.

The results of our algorithm were compared to results obtained by a non iterative triangulation scheme. Figure 6 presents the initial triangulation computed from the automatically extracted points and the associated reconstructed right image. In that algorithm homography estimation has been processed only with the 3 vertices and the fundamental matrix. Figure 7 shows the final triangulation and the resulting reconstructed right image. Comparing the two right reconstructed images leads to validate our approach. To access quality, note the reconstruction on straight lines (tables and posters edges) and on the dotted regular pattern.

5 Conclusion

We have presented a planar facet segmentation algorithm with some results which prove the benefit of our method. The dense field estimation can also be used separately to make a dense geometrically consistent 3D reconstruction while this resulting motion field exactly verify the epipolar constraint (i.e. the perspective projection model). Our global algorithm is able to deal with complex scenes without human interactions. Our future work will consist in self-calibrating our sequence in order to compute an Euclidean 3D model of the scene suitable to be visualized as a VRML representation.

References

[1] M. Black and P. Anandan. Robust incremental optic flow. In *Proceedings of the Conf. on Computer Vision and Pattern Recognition*, 1992.

[2] J. Blanc and R. Mohr. Towards fast and realistic image synthesis from real views. In *Scandinavian Conference on Image Analysis*, Finland, 1997.

[3] B. Boufama and R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. In *Proceedings of the International Conf. on Computer Vision*, pages 1030–1036, Cambridge, Massachusetts, 1995.

[4] S. Bougnoux and L. Robert. Totalcalib: a fast and reliable system for off-line calibration of images sequences. In *Proceedings of the Conf. on Computer Vision and Pattern Recognition*, 1997.

[5] B. Couapel and K. Bainian. Stereo vision with the use of a virtual plane in the space. *Chinese Journal of Electronics*, 4(2):32–39, 1995.

[6] G. Lemestre and D. Pelé. Trinocular image analysis for virtual fram reconstruction. In *Proceedings of SPIE VCIP*, San Jose, California, 1996.

[7] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. To appear in IEEE trans. I.P., available at ftp://ftp.irisa.fr/techreports/1996/PI-991, 1998.

[8] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the 6th International Conf. on Computer Vision*, pages 90–95, 1998.

[9] L. Robert and R. Deriche. Dense depth map reconstruction using a multiscale regularization approach which preserves discontinuities. In *Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging*, pages 32–39, Septembre 1995.

[10] L. Robert and O. Faugeras. Relative 3-D positioning and 3-D convex hull computation from a weakly calibrated stereo pair. *Image and Vision Computing*, 13(3):189–197, 1995.

[11] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78(1-2):87–119, 1994.

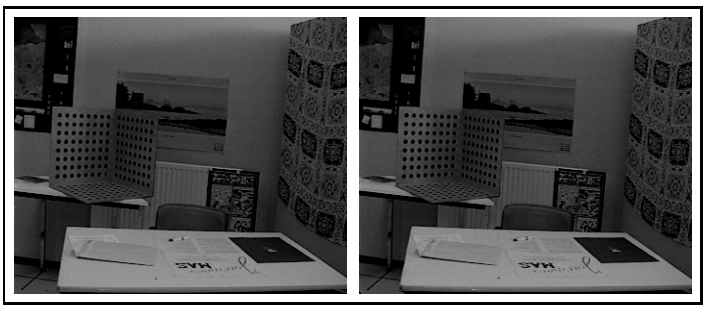

Figure 4: original left and right images

Figure 5: a,b: discontinuity and observation map (the darker, the more limited the energy contribution) - c: final disparity field

Figure 6: a: initial triangulation - b: initial reconstructed right image -c: resulting difference image

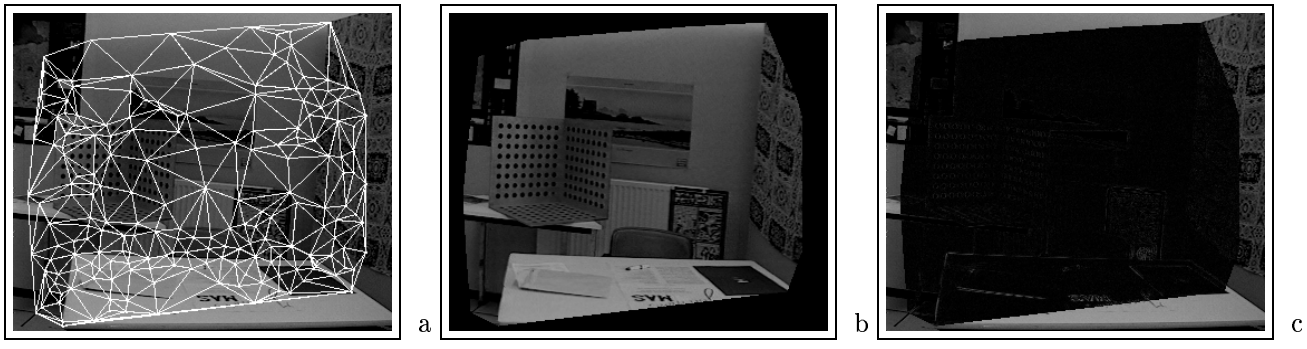

Figure 7: a: final triangulation - b: final reconstructed right image - c: resulting difference image