Constrained Structure and Motion Estimation from Optical Flow

Marco Zucchelli[†]   José Santos-Victor[‡]   Henrik I. Christensen[†]

†CVAP & CAS, Royal Institute of Technology, Stockholm, Sweden S100 44
‡Vislab, Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisboa, Portugal
zucch@nada.kth.se   —   jasv@isr.ist.utl.pt   —   hic@nada.kth.se

Abstract

Unbiased and consistent estimates of structure and motion can be obtained by least squares minimization of the differential epipolar constraint. Previous work on this subject does not make use of geometrical constraints that often are present in natural and man built scenes. This paper shows how linear constraints among feature points (collinearity and coplanarity) and bilinear relations among such entities (parallelism and incidence) can be incorporated in the minimization process to improve the structure and motion estimates. There are 2 main contributions: (i) the formulation of a constrained minimization problem for structure and motion estimation from optical flow and (ii) the solution of the optimization problem by Levenberg-Marquardt *and* direct projection. *We show that the proposed approach is stable, fast and efficient.*

1. Introduction

Optical flow has been efficiently used for structure and motion estimation. The differential epipolar constraint has been manipulated in different way in order to linearize the minimization problem [1, 2] . Although fast and of closed form, linear techniques are shown to lead to biased estimation of motion and structure. Zhang and Tomasi [3] recently demonstrated that such bias is due to the incorrect choice of the objective function and that unbiased and consistent estimates can be obtained by minimization of the differential epipolar constraint in the least squares sense. However, this approach does not exploit any constraints of the scene geometry. Man built and natural environments feature many geometrical entities (lines and planes for example) often arranged in special configurations (parallelism, incidence etc). Attempts to incorporate such constraints in reconstruction algorithms have been previously done in the stereo framework [4, 5, 6].

In this paper we approach the problem of using such geometrical constraints in the differential structure from motion setting. We show (i) how to incorporate linear constraints among feature points (collinearity and coplanarity)

and bilinear relationships among such entities (parallelism and incidence at a certain angle) in the minimization of the differential epipolar constraint; (ii) we show how fast and stable convergence is obtained by the use of a Levenberg-Marquardt iterative minimization and a *direct projection method* [7] for constraints enforcement. The result is a fast and efficient algorithm which drastically reduces the reconstruction error.

Note that in this paper we do not discuss the problem of *segmentation*: we assume that planes and lines are already detected by one of the existing segmentation techniques [8].

2. Unconstrained Non Linear Estimation of Structure and Motion

2.1   Problem Formulation

The relationship between the image plane motion field $\mathbf{u}(\mathbf{x})$ and the motion of the camera is expressed as :

$$\mathbf{u}(\mathbf{x}) = \frac{1}{Z}A(\mathbf{x})v + B(\mathbf{x})\omega + \mathbf{n}(\mathbf{x}) \qquad (1)$$

where (v, ω) are the linear and angular camera velocities and $\mathbf{n}(\mathbf{x}) \sim N(0, \sigma)$ is zero-mean gaussian additive noise. Z is the depth of the scene points whose 3D position we indicate with $\mathbf{X} = [X, Y, Z]$. The matrices $A(\mathbf{x})$ and $B(\mathbf{x})$ are functions of image coordinates defined as follows:

$$B = \begin{bmatrix} -xy & (1+x^2) & -y \\ -(1+y^2) & xy & x \end{bmatrix} ; A = \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix}$$

We assume that the focal length is equal to 1, so that the projection model is such that $\mathbf{x} = \frac{1}{Z}[X, Y]$.

2.2   Unconstrained Optimization

Given two views of the same scene, the instantaneous motion model of Eq.(1) is valid when the camera rotation is small and the forward translation is small relative to the depth. If this condition is met, optical flow between the two frames can be computed and depths and velocities can be

estimated. We define the residual for the i_{th} feature relative to the pair of as:

$$\mathbf{r}_i = \mathbf{u}_i - \frac{1}{Z_i} A(\mathbf{x}_i)v - B(\mathbf{x}_i)\omega \qquad (2)$$

We stack the residuals $\mathbf{r}_i$ in the $2N \times 1$ vector $\rho = [\mathbf{r}_1, \ldots, \mathbf{r}_N]$ and the structure and motion parameters in the vector $\theta = (v, \omega, \mathbf{Z})$ with $\mathbf{Z} = (Z_1, \ldots, Z_N)$. Structure and motion can be estimated by solving the least squares problem:

$$\hat{\theta} = arg \min_{\theta} \|\rho(\theta)\|^2 \qquad (3)$$

.

The problem in Equation (3) is a non linear least squares estimation and has to be solved by an iterative technique. We used Gauss-Newton in the form:

$$J_k \Delta\theta_k = -\rho_k \qquad (4)$$

where J is the Jacobian of ρ, and k is the iteration index.

Initialization: Iterative techniques for the solution of non linear optimization problems are locally convergent and a good initialization is needed in order to find the absolute minimum. Initialization is made easier in our problem due to the *separability* of the differential epipolar constraint equation. Defining:

$$\mathbf{e} = [e_1, e_2] = \frac{A(\mathbf{x})v}{\|A(\mathbf{x})v\|} \qquad (5)$$

the vector $\tilde{\mathbf{e}} = [e_1, -e_2]^T$ is normal to the component of the optical flow generated by the linear velocity:

$$\tilde{\mathbf{e}} \cdot \frac{1}{Z} A(\mathbf{x})v = 0 \qquad (6)$$

Multiplying Eq. (1) by $\tilde{\mathbf{e}}$ we get:

$$\tilde{\mathbf{e}}(\mathbf{u} - B(\mathbf{x})\omega) = 0 \qquad (7)$$

from which ω can be estimated by least squares minimization when v is known. When v and ω are known we can estimate the ratio $\frac{1}{Z}$ from:

$$\frac{1}{Z} = \frac{\mathbf{e}^T(\mathbf{u} - B(\mathbf{x})\omega)}{\|A(\mathbf{x})v\|} \qquad (8)$$

To generate the initial value for the variables $(v, \omega, \mathbf{Z})$ it is sufficient to initialize the vector v on the half sphere of radius 1 and then estimate the respective ω and $\mathbf{Z}$ using Eq. (7) and Eq. (8).

3. Constrained Non Linear Estimation of Structure and Motion

3.1 Constraints Formulation

We consider only constraints linear in the features coordinates (i.e. collinearity and coplanarity) and bilinear constraints among such geometrical entities (i.e. parallelism

and incidence at a certain angle). Planes are parameterized by a direction $\mathbf{p}$ and a distance d. Lines are described by the intersection of 2 planes. Constraints can be expressed as:

$$\mathbf{X}\mathbf{p}_a = d_a \qquad\qquad \mathbf{X} \in plane \qquad (9)$$

$$\mathbf{p}_a\mathbf{p}_b = \theta_{a,b} \qquad incidence\ at\ angle\ \theta_{h,l} \qquad (10)$$

$$\mathbf{p}_a = \mathbf{p}_b \qquad\qquad parallelism \qquad (11)$$

3.2 Constrained Optimization

Coplanarity and collinearity can be easily incorporated in equation Eq. (3) by parameterization. If a set of points $\mathbf{X}_h$, $h \in \{1 \ldots H\}$ belongs to a plane, their depth can be parameterized as $Z_h = \frac{d}{[x_h, y_h, 1]\mathbf{p}}$. The residual of such points takes the form:

$$\mathbf{r}_h = \mathbf{u}_h - A(\mathbf{x}_h)\frac{[x_h, y_h, 1]\mathbf{p}}{d}v - B(\mathbf{x}_h)\omega \qquad (12)$$

Indicating with $\mathbf{c}$ the vector of the parameters of the M constraints, i.e. $\mathbf{c} = [\mathbf{p}_1, d_1 \ldots \mathbf{p}_M, d_M]$, structure and motion can be estimated by solving:

$$\hat{\Theta} = arg \min_{\boldsymbol{\theta}} \|\rho(\Theta)\|^2 \qquad (13)$$

where $\Theta = (v, \omega, \mathbf{Z}_u, \mathbf{c})$ and $\mathbf{Z}_u$ is the vector of the depths of the points for which no constraints are available . To speed up the solution we used Levenberg-Marquardt iteration technique instead of Gauss-Newton:

$$\mathbf{L}_k \Delta\Theta_k = -\mathbf{J}_k^T \rho_k \qquad (14)$$

where $\mathbf{J}$ is the jacobian of $\rho(\Theta)$. The matrix $\mathbf{L}_{m,n}$ is defined as $\mathbf{L}_{m,m} = (1+\lambda)(\mathbf{J}^T\mathbf{J})_{m,m}$ and $\mathbf{L}_{m,n} = (\mathbf{J}^T\mathbf{J})_{m,n}$ for $m \neq n$. The parameter λ is fixed to be 10^{-3} at the beginning.

Constraints among geometrical entities (Eq. (10) and (11)) define a feasible space where to search to the absolute minimum of the function we are optimizing. To solve the constrained minimization problem we used a *direct projection* method [7]: constraints of the form $f(\mathbf{c}) = 0$ are differentiated obtaining linear equations in $\Delta\mathbf{c}$:

$$\frac{\partial f}{\partial \mathbf{c}}\Delta\mathbf{c} = 0 \qquad (15)$$

Such equations are added to the system in Eq. (14) which is solved as for normal Levenberg-Marquardt iterations. Due to the fact that the constraints are replaced by their differentials, the incremented structure and motion parameters Θ_{k+1} do not belong to the feasible space but to its tangent space. To keep the solution in the feasible space the vector Θ_{k+1} is re-projected onto it, i.e. $\Theta_{k+1} \mapsto \mathcal{P}(\Theta_{k+1})$ where $\mathcal{P}$ is the projector.
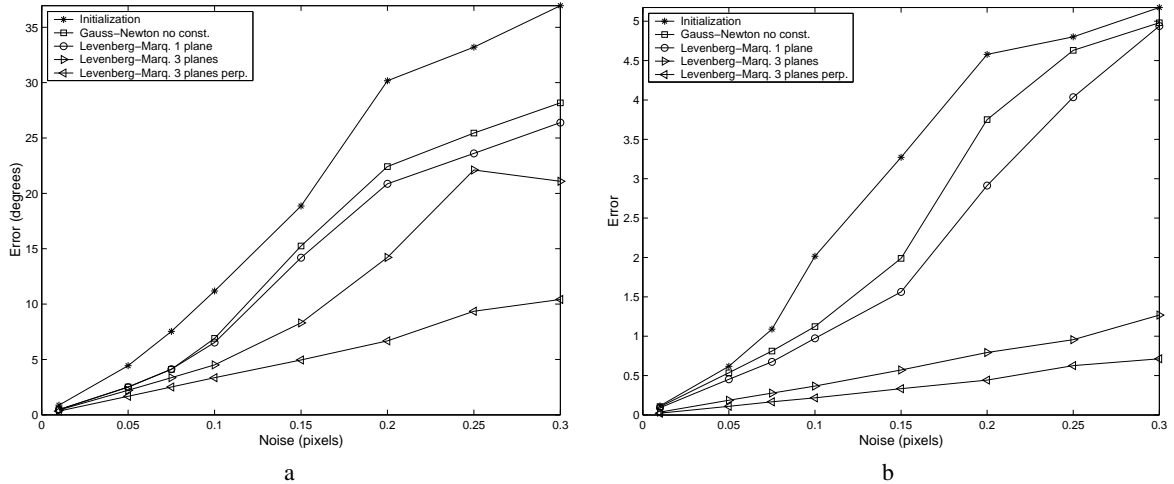
Figure 1. (a) Linear velocity error. (b) Structure error.

This approach is known as *direct projection*. The main advantage of this technique is that a constrained minimization problem is transformed into an un-constrained one that can be solved by a standard iterative technique. The drawback is that convergence is not guaranteed since the minimization method finds a descent direction in the tangent space which does not ensures that such direction projected onto the feasible space is still descent. Formally this means that $\|\rho(\Theta_{k+1})\|^2 < \|\rho(\Theta_k)\|^2$ that does not imply that $\|\rho(\mathcal{P}(\Theta_{k+1})\|^2 < \|\rho(\Theta_k)\|^2$. Nevertheless for the class of problems we are trying to solve this approach has very good convergence properties. This are reviewed more in detail in the experiments section.

Initialization For constrained minimization parameters for planes and lines must also be initialized. We first estimate an initial structure and motion as in the unconstrained case. The set of parameters **c** can be initially estimated by using the initial structure and Eq. (9). Such first estimate must be refined in order to have **c** in the feasible space. This is simply done by re-projection. For example, if M perpendicular planes are present the projector is the unitary transformation that align them with M perpendicular unitary vectors. The optimal solution to this problem can be easily computed by SVD (see [9]).

4. Results

We extensively tested the algorithm using synthetically generated flow fields. For homogeneity and simplicity we used the same experimental conditions and benchmarks as in [10]. The focal length of the simulated camera is set to 1 and the focal plane dimensions to 512×512 pixels. The field of view is 90 degrees. Random cloud of 100 points

are generated in a depth range of 2-8 focal lengths. The motion is a combination of rotations and translations. The rotational speed magnitude was constant and chosen to be 0.23 degrees per frame. The magnitude of the linear velocity was chosen to fixate the point at the center of the random cloud. The average optical flow generate by this setup is approximately 1 pixel per frame which is very similar to real working conditions. Zero-mean gaussian noise of different std was added to the components of the velocity. The noise range was between 0 and 0.3 pixels which is a realistic estimate for our working conditions (see [11] and [10]).

Simulations We generated 48 points distributed on 3 orthogonal surfaces for testing purposes. Initialization was obtained by estimating the linear velocity by subspace methods [12] and then using Equations (7) and (8) to estimate ω and **Z**. We did 4 tests to asses the performance of the constrained minimization: we first used plane Gauss-Newton with no constraints, we then used Gauss-Newton with one plane constraint, three planes constraints and finally three orthogonal plane constraints. Figures 1 (*a*) and (*b*) show the linear velocity and structure errors for different noise levels. The error functions are described in the appendix.

Real Images Optical flow is estimated at sparse positions over the image plane by the method described in [13]. The average disparity for the two sequences used is approximately 2 pixels per frame. The total number of features used is 245 the Teabox and 271 for the Linuxbox sequence.

Convergence Properties To speed up the convergence the Levenberg-Marquardt iterative algorithm was used. For a noise of 0.1 pixels, which is the normal amount expected in our working conditions (see [11]), the global minimum is found essentially all the times within 10-15 steps. For very noisy fields, i.e. 0.3 pixels, the convergence rate is

about 70%. Convergence rate for noisy fields can be further increased by randomly initializing several times the algorithm and then taking the solution that generates the minimal residual flow. This process is made easy by the variable separability described in section 2.

5. Conclusions

In this paper we showed how linear geometrical constraints (collinearity and coplanarity) and bilinear constraints among them (incidence and parallelism) can be used to improve structure and motion estimation in the differential setting. While Gauss-Newton is proved to converge fast in the unconstrained case, Levbenberg-Marquardt has to be used when constraints are incorporated. The method of *direct projection*, used to enforce the constraints, is proved to be stable and efficient. Simulations and tests on real images show the efficiency of the proposed method.

Appendix A

Sensitivity in the 3D reconstruction is measured aligning the ground-truth and the reconstructed model and taking the average of the distance between estimated and true features positions, $\hat{\mathbf{X}}$ and $\mathbf{X}$, divided by the average object size. A similar approach is taken to assess the error in the reconstructed velocities:

$$\sigma_r = E[\frac{\|\mathbf{X}_i - \hat{\mathbf{X}}_i\|}{obj.\ size}]; \quad \sigma_v = \sqrt{\frac{1}{L-1}\sum_{l=1}^{L}[\cos^{-1}(\bar{v}\hat{v}_l)]^2}$$
(16)

where i runs over the features and $\bar{v}$ is the average of the reconstructed velocities, that minimizes $\sum_{l=1}^{L}\cos^{-1}\hat{v}_l\bar{v}$ subject to $\|\bar{v}\| = 1$. L is the number of trials.

The average rotation matrix $\bar{R}$ is computed from the estimates, $\hat{R}_l$. The rotation sensitivity is computed as the standard deviation of the difference angle, ϕ, between $\bar{R}$ and the estimates, $\hat{R}$, for each trial sample:

$$\sigma_\phi = \sqrt{\frac{1}{L-1}\sum_{l=1}^{M}\phi_l^2} \ ; \ \phi_l = \cos^{-1}[\frac{Tr(\hat{R}_l\bar{R})-1}{2}]$$
(17)

Figure 2. 2 views of a reconstructed model.
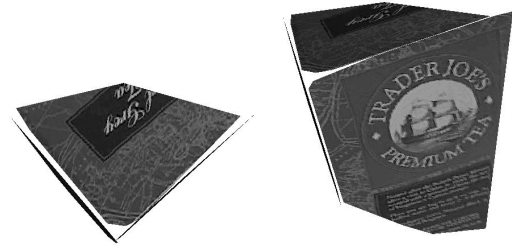
Figure 3. 2 views of a reconstructed model.

References

[1] Heeger D. J. and Jepson A. D. Subspace methods for recovering rigid motion. *IJCV*, 7(2):95–117, 1992.

[2] Kosecka J. Ma Y. and Sastry S. Linear differential algorithm for motion recovery:a geometric approach. *IJCV*, 36(1):71–89, 2000.
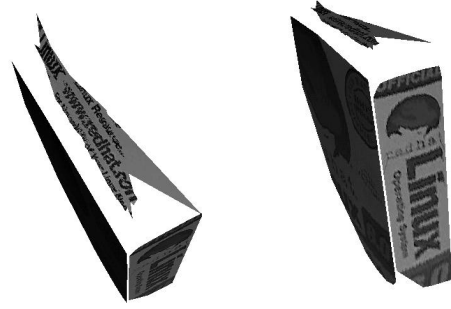
[3] Zhang T. and Tomasi C. Fast, robust, and consistent camera motion estimation. In *CVPR*, volume 1, pages 164–170, 1999.

[4] Grossman E. and Santos-Victor J.A. Dual representations for vision-based 3d reconstruction. In *BMVC*, pages 516–525, 2000.

[5] Szeliski R. and Torr P.H. Geometrically constrained structure from motion: Points on planes. In *SMILE workshop*, pages 171–186, 1998.

[6] Bondyfalat Didier and Bougnoux S. Imposing euclidean constraint during self-calibration processes. In *SMILE Workshop*, pages 224–235, 1998.

[7] Scales L.E. *Introduction to Non-Linear Optimization*. Macmillan, 1985.

[8] Haralick R.M. and Shapiro L.G. Image segmentation techniques. *CVGIP*, 29(1):100–132, 1985.

[9] Kanatani K. *Geometric Computation for Machine Vision*. Clarendon Press, 1993.

[10] Tomasi C. and Heeger D. J. Comparison of approaches to egomotion computation. In *CVPR*, pages 315–320, 1994.

[11] Fleet D. J. Barron J. L. and Beuchemein S. S. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[12] MacLean W. J. Removal of translational bias when using subspace methods. In *ICCV*, pages 753–758, 1999.

[13] Shi J. and Tomasi C. Good features to track. In *CVPR*, pages 593–600, 1994.