Motion estimation using the differential epipolar equation

L. Baumela
Dep. Inteligencia Artificial
U. Politécnica de Madrid
28660 Madrid (Spain)
lbaumela@fi.upm.es

L. Agapito
Dep. of Engineering Science
University of Oxford
19 Parks Rd, Oxford (UK)
lourdes@robots.ox.ac.uk

P. Bustos
Dep. Informática
U. Extremadura
Cáceres (Spain)
pbustos@unex.es

I. Reid
Dep. of Engineering Science
University of Oxford
19 Parks Rd, Oxford (UK)
ian@robots.ox.ac.uk

Abstract

We consider the motion estimation problem in the case of very closely spaced views. We revisit the differential epipolar equation providing an interpretation of it. On the basis of this interpretation we introduce a cost function to estimate the parameters of the differential epipolar equation, which enables us to compute the camera extrinsics and some intrinsics. In the synthetic tests performed we compare this continuous method with traditional discrete motion estimation and contrary to previous findings [6] cannot perceive any computational advantage for the continuous method.

1. Introduction

Consider a camera with unknown (and possibly varying) intrinsic parameters moving in a static scene. At any time instant, the projection of the scene onto the image plane generates an *optical flow field*. Each flow vector of this field has a position component $\mathbf{m} = [u, v, 1]^\top$ and a velocity component $\dot{\mathbf{m}} = [\dot{u}, \dot{v}, 0]^\top$, where u and v are the coordinates of the projection of a world point onto the image plane, and $(\dot{u}, \dot{v})$ is the instantaneous velocity vector.

The *differential epipolar equation* [2, 5]

$$\mathbf{m}^\top \mathbf{V} \dot{\mathbf{m}} + \mathbf{m}^\top \mathbf{C} \mathbf{m} = 0 \tag{1}$$

with $\mathbf{C}$ symmetric and $\mathbf{V}$ anti-symmetric, is the first order relation between the optical flow field and the camera motion and intrinsic parameters.

In this paper we estimate the camera motion parameters using the differential epipolar equation by minimising a meaningful geometric error function. We then compare this to the motion estimates obtained from the more common discrete epipolar equation. Sadly – and contrary to the findings reported in [6] – we find no clear evidence that any performance benefits result.

We begin by reviewing the differential epipolar equation in §2, discuss our implementation of the motion estimation problem in §3, and in §4 show experimental results, comparing these to the current state-of-the-art in discrete motion estimation. We draw conclusions in §5.

2. The differential epipolar equation

In our present work, we are considering the case of very closely spaced views. As indicated in the previous section, the geometry of this case is governed by the *differential epipolar constraint* (1) relating image points and their instantaneous image velocities, which was introduced in [5]. In [1], this was extended to a general set of differential constraints (relating image points and their instantaneous velocities, accelerations, etc) analogous to the multiview constaints described in [7, 4].

The two view constraint in that work comes from a first order taylor expansion of the projection equation $\lambda(t)\mathbf{m}(t) = \mathbf{P}(t)\mathbf{X} = [\mathbf{Q}(t) \mid T(t)]\,\mathbf{X}$, viz:

$$\begin{aligned} & (\lambda + \dot{\lambda}t + O(t^2))(\mathbf{m} + \dot{\mathbf{m}}t + O(t^2)) \\ = \;& ([\mathbf{Q} \mid T] + [\dot{\mathbf{Q}} \mid \dot{T}]t + O(t^2))\mathbf{X} \end{aligned} \tag{2}$$

If we choose a coordinate frame (as we are free to do) which aligns the world frame with the camera, and rewrite the equation in matrix form we obtain:

$$\begin{bmatrix} \mathbf{I} & 0 & \mathbf{m} & 0 \\ \dot{\mathbf{Q}} & \dot{T} & \mathbf{m} & \dot{\mathbf{m}} \end{bmatrix} \begin{bmatrix} -\mathbf{X} \\ \lambda \\ \dot{\lambda} \end{bmatrix} = \mathbf{0} \tag{3}$$

The 6x6 matrix on the left has a non-zero null space, hence its determinant is zero. Laplacian determinant expansion then yields the equation

$$m^i \dot{m}^k \epsilon_{ijk} \dot{t}^j - m^i m^l \epsilon_{ijk} \dot{t}^j \dot{q}^k_l = 0 \tag{4}$$

which is exactly (1) but restated in tensor notation with $\dot{T}$ and $\dot{\mathbf{Q}}$ represented respectively as $\dot{t}^j$ and $\dot{q}^k_l$.

Although this derivation is very clean and direct, it hides some interesting structure which we would like to explore further, and so we provide a derivation which is more in the spirit of [5, 6] as follows.

The world point $\mathbf{X}$ projects to the two image points as

$$z_1\mathbf{m}_1 = \mathbf{K}_1[\,\mathbf{I}\,|\,\mathbf{0}\,]\,\mathbf{X} \quad (5)$$

$$z_2\mathbf{m}_2 = \mathbf{K}_2[\,\mathbf{R}\,|\mathbf{t}\,]\,\mathbf{X} \quad (6)$$

where the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t} = (t_x, t_y, t_z)^\top$ represent camera motion, $\mathbf{K}_1, \mathbf{K}_2$ camera intrinsics and z_1, z_2 relative depths caused by motion.

Combining (5) and (6) we get

$$z_2\mathbf{K}_2^{-1}\mathbf{m}_2 = z_1\mathbf{R}\mathbf{K}_1^{-1}\mathbf{m}_1 + \mathbf{t}. \quad (7)$$

from which the derivation of the familiar uncalibrated epipolar equation is relatively straighforward.

Our goal here is to study this relation in the continuous case. That is, we will assume that $t_2 = t_1 + dt$ and compute the limit of (7) when $dt \to 0$. We will consider small variations of:

- Intrinsics $\mathbf{K}_2^{-1} = \mathbf{K}_1^{-1} + \dot{\mathbf{K}}^{-1}dt + o(dt^2)$.
- Image plane motion: $\mathbf{m}_2 = \mathbf{m}_1 + \dot{\mathbf{m}}\,dt + o(dt^2)$.
- Camera rotation: $\mathbf{R} = \mathbf{I} + \mathbf{\Omega}\,dt + o(dt^2)$.
- Camera translation: $\mathbf{t} = \mathbf{v}\,dt + o(dt^2)$.
- Depth: $z_2 = z_1 + \dot{z}\,dt + o(dt^2)$.

Introducing the previous equations into (7) and taking the limit when $dt \to 0$ we obtain the explicit formula for $\dot{\mathbf{m}}$ when the camera is translating and rotating simultaneously

$$\dot{\mathbf{m}} = \underbrace{[-\mathbf{K}_1\dot{\mathbf{K}}^{-1} + \mathbf{K}_1\mathbf{\Omega}\mathbf{K}_1^{-1}]}_{\dot{\mathbf{H}}_\infty}\mathbf{m}_1 - \underbrace{\frac{\dot{z}}{z_1}}_{1/\tau}\mathbf{m}_1 + \frac{1}{z_1}\underbrace{\mathbf{K}_1\mathbf{v}}_{e_3\mathbf{e}} \quad (8)$$

where $e_3\mathbf{e}$ is the focus of expansion (or instantaneous epipole) of the image, τ is the time to contact z_1 and $\dot{\mathbf{H}}_\infty$ is the derivative of the homography for the plane at infinity. Rewriting (8) we get

$$\dot{\mathbf{m}} = \underbrace{\left(\dot{\mathbf{H}}_\infty - \mathbf{s}^\top\dot{\mathbf{H}}_\infty\mathbf{m}_1\mathbf{I}\right)\mathbf{m}_1}_{\dot{\mathbf{m}}_\Omega} + \underbrace{\frac{e_3}{z_1}(\mathbf{e} - \mathbf{m}_1)}_{\dot{\mathbf{m}}_v} \quad (9)$$

where $\mathbf{s}^\top = (0, 0, 1)$, $\dot{\mathbf{m}}_v$ is the component of optical flow caused by camera translation and $\dot{\mathbf{m}}_\Omega$ is the component caused by camera rotation and intrinsics variation. Here we can observe that the instantaneous motion of an image point (optical flow) caused by camera rotation and translation with varying intrinsics is the composition of two flows: a rotation/intrinsic flow, $\dot{\mathbf{m}}_\Omega$, whose value can be computed, provided the derivative of the infinite homography is known, and a translational flow, $\dot{\mathbf{m}}_v$, whose value can only be computed up to a scale factor, because it depends on (e_3), the translational component along the optical axis.

Taking the dot product of (8) with $\mathbf{m}_1^\top[\mathbf{K}_1\mathbf{v}]_\times$ we get the differential epipolar equation

$$\mathbf{m}_1^\top\underbrace{[\mathbf{K}_1\mathbf{v}]_\times}_{\mathbf{V}=[\mathbf{e}]_\times}\dot{\mathbf{m}} + \mathbf{m}_1^\top\underbrace{[\mathbf{K}_1\mathbf{v}]_\times[-\mathbf{K}_1\dot{\mathbf{K}}^{-1} + \mathbf{K}_1\mathbf{\Omega}\mathbf{K}_1^{-1}]}_{\mathbf{C}=[\mathbf{e}]_\times\dot{\mathbf{H}}_\infty}\mathbf{m}_1 = 0. \quad (10)$$

From (10) it follows that $\mathbf{V}$ is an antisymmetric matrix encoding the projective coordinates of the instantaneous epipole (FOE), the estimated $\mathbf{C}$ is the symmetric part of $[\mathbf{e}]_\times\dot{\mathbf{H}}_\infty$, and consequently that the constraint $\mathbf{e}^\top\mathbf{C}\mathbf{e} = 0$ must hold.

We will use equation (10) to self calibrate a moving camera with varying focal length. This problem was previously addressed in [8] for a moving camera with fixed and unknown intrinsics using normal flow measurements.

3. Estimation of the differential epipolar equation parameters

In this section we will present a procedure to estimate the parameters of the epipolar equation. It is based on the assumption that our optical flow estimation algorithm provides exact values for $\mathbf{m}$ and approximated (or noisy) values for $\dot{\mathbf{m}}$.

First we will introduce a cost function that measures the extent to which optical flow data satisfy (1) for a given set of parameters $\mathbf{C}$, $\mathbf{V}$. Then, we'll introduce the constraint on the estimated parameters and resort to an iterative minimisation procedure to compute the optimal values for $\mathbf{C}$ and $\mathbf{V}$.

Let $c_1, c_2, c_3, c_4, c_5, c_6$ be the components of Sym($\mathbf{C}$) and $\mathbf{e}^\top = (e_1, e_2, e_3)$. Substituting these values for $\mathbf{V}$ and $\mathbf{C}$ into (1) and factoring the noisy terms $\dot{u}$, $\dot{v}$, we get

$$\dot{u}\underbrace{(e_3v - e_2)}_{p} + \dot{v}\underbrace{(e_1 - e_3u)}_{q} + \underbrace{c_1u^2 + 2c_2uv + 2c_3u + c_4v^2 + 2c_5v + c_6}_{r} = 0. \quad (11)$$

The line $\Lambda_i(\dot{u}_i, \dot{v}_i) \equiv p\,\dot{u}_i + q\,\dot{v}_i + r = 0$ represents the *instantaneous epipolar line* for the point (u_i, v_i). From (9) it is easy to see that it is parallel to the line joining $\mathbf{m}$ and $\mathbf{e}$ ($\overline{\mathbf{me}}$), its position depending on the value of $\dot{\mathbf{H}}_\infty$. When the motion is only translational and intrinsics are fixed ($\dot{\mathbf{H}}_\infty = \mathbf{0}$) it coincides with $\overline{\mathbf{me}}$ (see Fig 1).

Note that normal components of optical flow do not give us any direct constraints on the epipolar geometry. Hence, in the absence of other constraints (such as scene planarity), we are limited to using those flow vectors obtained from regions with low autocorrelation in all spatial directions – i.e. at 'corners'.

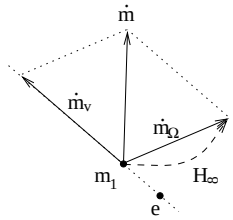

Figure 1. Interpretation of Λ_i.

Our goal is to find the parameters of $\mathbf{V}$ and $\mathbf{C}$ that minimise $d(\dot{\mathbf{m}}_i, \Lambda_i)$, the geometric distance of $\dot{\mathbf{m}}_i$ to its instantaneous epipolar line Λ_i. The cost function used, J, is the sum of the squared distances of each point:

$$J(\mathcal{O}, \theta) = \sum_{i=1}^{N} d^2(\dot{\mathbf{m}}_i, \Lambda_i) = \sum_{i=1}^{N} \frac{(p_i\dot{u}_i + q_i\dot{v}_i + r_i)^2}{p_i^2 + q_i^2},$$

where $\mathcal{O} = \mathbf{m}_1, \dot{\mathbf{m}}_1, \ldots, \mathbf{m}_N, \dot{\mathbf{m}}_N$ is the optical flow data and $\theta = (e_1, e_2, e_3, c_1, c_2, c_3, c_4, c_5, c_6)^\top$ is the parameter vector.

The uncertainty of each velocity measurement $(\dot{u}, \dot{v})$ can be described by a 2×2 covariance matrix, $\mathbf{\Sigma}_{\dot{\mathbf{m}}}$. We then adopt the following cost function, which takes into account this information

$$J(\mathcal{O}, \theta) = \sum_{i=1}^{N} \frac{\theta^\top \mathbf{f}_i \mathbf{f}_i^\top \theta}{\mathbf{e}^\top \mathbf{H}_i \mathbf{\Sigma}_{\dot{\mathbf{m}}_i} \mathbf{H}_i^\top \mathbf{e}}, \tag{12}$$

where

$$\mathbf{f}_i = (\dot{v}_i, -\dot{u}_i, \dot{v}_i u_i - \dot{u}_i v_i, u_i^2, 2u_i v_i, 2u_i, v_i^2, 2v_i, 1)^\top \tag{13}$$

and

$$\mathbf{H}_i^\top = \begin{pmatrix} 1 & 0 & -v_i \\ 0 & -1 & u_i \end{pmatrix}.$$

We then require the minimum of J subject to the constraint $\mathbf{e}^\top \mathbf{C} \mathbf{e} = 0$. In our solution we substitute one of the θ parameters in (12) for the explicit value obtained from the constraint and employ a general (unconstrained) iterative minimisation procedure. For example,

$$c_6 = -\frac{c_1 e_1^2 + 2c_2 e_1 e_2 + c_4 e_2^2}{e_3^2} - \frac{2c_3 e_1 + 2c_5 e_2}{e_3}. \tag{14}$$

By substituting c_6 in (12) by (14) we get a new cost function J_c that implictly imposes the algebraic constraint

$$J_c(\mathcal{O}, \delta) = \sum_{i=1}^{N} \frac{\delta^\top \mathbf{g}_i \mathbf{g}_i^\top \delta}{\mathbf{e}^\top \mathbf{H}_i \mathbf{\Sigma}_{\dot{\mathbf{m}}_i} \mathbf{H}_i^\top \mathbf{e}}, \tag{15}$$

where $\delta = (e_1, e_2, e_3, c_1, c_2, c_3, c_4, c_5)^\top$ and

$$\mathbf{g}_i^\top = \Big[\dot{v}_i, \dot{u}_i, \dot{u}_i v_i - \dot{v}_i u_i, u_i^2 - \frac{e_1^2}{e_3^2}, 2(u_i v_i - \frac{e_1 e_2}{e_3^2}), 2(u_i - \frac{e_1}{e_3}), v_i^2 - \frac{e_2^2}{e_3^2}, 2(v_i - \frac{e_2}{e_3})\Big].$$

Once the epipolar equation's parameters are estimated, computing the FOE is straightforward from (10). The rotational velocity of the camera, Ω, and the focal length can also be computed given partial knowledge of the camera calibration parameters [2].

4. Experimental results

The experimental results reported here are synthetic tests performed with data obtained from a simulated camera with realistic parameters. In each of the tests performed a cloud of 400 points is randomly generated in a cube of 5 meters of depth located at a distance of 2.5 meters and centered in front of the camera. Each point is projected onto the image plane and associated to each point we compute an instantaneous flow vector, which is the projection onto the image of the instantaneous velocity of the point relative to the camera. Horizontal and vertical components of the flow vector are contaminated with randomly distributed gaussian noise. Each of the following test is repeated 20 times and the data reported are the average of the computed results.

Our main goal in this paper is to compare the precision in the estimation of the FOE of the continuous model presented here with traditional discrete methods. Here the fundamental matrix is determined using non-linear estimation which enforces the rank-2 constraint, and the epipole as the right null space of $\mathbf{F}$. In the first test (see Fig.2) we keep the camera parameters fixed and increase the contaminating noise. In the second test (see Fig.3) noise remains fixed with varying rotational and translational disparity. In both plots continuous model results are represented with solid line and the discrete ones with dashed line.

Contrary to the experimental results reported in [6], we do not perceive in the synthetic tests a clear advantage of one method over the other.

5. Conclusions

We have derived the continuous analogue of the discrete epipolar equation, given a geometric interpretation of it, and a practical algorithm for computing a camera's motion parameters from closely spaced views.

The results, when compared with the current state-of-the-art in discrete epipolar geometry, fare no better, and in some cases worse, than the discrete version. Triggs [3] has reported (although not demonstrated) a similar conclusion.

The theoretical studies in [5, 2, 1, 3] are of some value in furthering our understanding of geometric constraints.

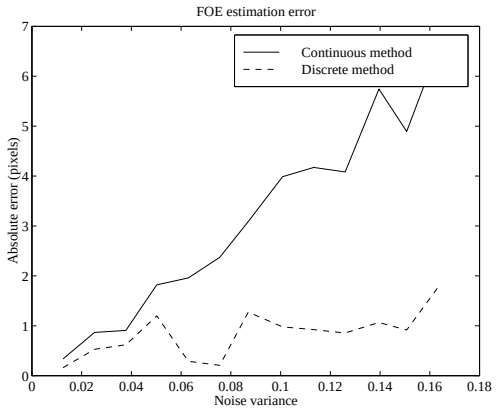

Figure 2. FOE estimation noise varying.

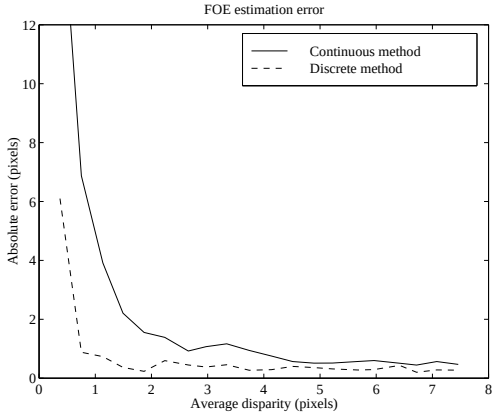

Figure 3. FOE estimation disparity varying.

Triggs forwards two potential practical advantages of the differential approach: (i) the correspondence problem is easier; (ii) the problem may be less non-linear. Unfortunately, as our results show, the former would not seem to be an advantage given that discrete approaches appear to perform as well as continuous, even for closely spaced views, and – at any rate in our formulation – the latter is not because of the need to impose the algebraic constraint.

It may be possible to improve the results somewhat by considering long sequences, however in these circumstances one would expect discrete methods to improve further since usually a longer sequence will correspond to a larger viewing baseline. Although our results are not totally conclusive, they do seem to suggest that there is doubt as to whether or not further practical investigation will yield practical advantages of the continuous method.

A second possibility is to investigate the continuous analogue of the tri-focal tensor. [1] did this in implicit form and we have performed a determinant expansion of their constraint to obtain four tensors which encode the relative geometry of points and their image velocities and accelerations. Two practical problems arise on further investigation. The first is the unweildy number of constraints which exist between the tensor elements (and how best to enforce them), and the second is the perrenial one of how best to obtain accurate derivatives (especially higher order ones) from noisy image data.

Acknowledgements

The authors gratefully acknowledge the support of an Acciones Integradas grant from the Spanish Ministry of Education and The British Council.

References

[1] K. Åström, A. Heyden. “Continuous time matching constraints for image streams.” *Int’l J. of Computer Vision*. Vol. 28, N. 1, pp. 85-96. 1998.

[2] M.J. Brooks, W. Chojnacki, L. Baumela. “Determining the egomotion of an uncalibrated camera from instantaneous optical flow.” *J. of the Optical Society of America A.* Vol. 14, N. 10, pp. 2670-2677. 1997.

[3] W. Triggs. “Differential Matching Constraints”. *Proc. 7th Int’l Conf. on Computer Vision*. pp. 370-376. 1999.

[4] W. Triggs. “The geometry of projective reconstruction i: Matching constraints and the joint image”. *Proc. 5th Int’l Conf. on Computer Vision*. pp. 338-343. 1995.

[5] T. Viéville, O.D. Faugeras. “Motion analysis with a camera with unknown and possibly varying parameters.” *Proc. 5th Int’l Conf. on Computer Vision.* pp.750-756. 1995.

[6] T. Viéville, O.D. Faugeras. “The first order expansion of motion equations in the uncalibrated case.” *CVGIP: Image Understanding*. Vol. 64. N. 1. pp. 128-146. 1996.

[7] O.D. Faugeras, B. Mourrain. ”On the geometry and algebra of point and line correspondences between N images.” *Proc. 5th Int’l Conf. on Computer Vision*. pp. 951-962. 1995.

[8] T. Brodský, C. Fermüller, Y. Aloimonos. “Self-Calibration from image derivatives.” *Proc. Int’l Conf. on Computer Vision*. pp. 83-89. 1998.