14th Intern. Conf. on Pattern Recognition — ICPR'98 Australia

Egomotion Estimation on a Topological Space

César Silva* José Santos-Victor
Instituto de Sistemas e Robótica - Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL
e-mail: {cesar,jasv}@isr.ist.utl.pt

Abstract

We present an egomotion estimation method for a monocular observer moving with arbitrary motion. The method is uniquely based on the normal flow field and the estimation problem is defined in a particular topological space — the $\mathcal{L}$-space — to allow the use of global data for the final estimates. Additionally, robust statistics methods are used for motion model selection, and provide improved robustness to image noise and data outliers.

1. Introduction

We address the problem of egomotion estimation for a a monocular observer with arbitrary translation and rotation. This problem has been approached before either by using point correspondences between various frames [4] or by computing the dense optical flow field [2]. Alternatively, direct methods rely exclusively on the normal flow field[1], thus overcoming the ill-posedness nature of the flow estimation image matching problems.

We extend previous work [6], by formulating the problem on a different space — **the $\mathcal{L}$-space** — that allows for numerous normal flow vectors to contribute for the problem solution, hence improving robustness and reliability. The input data consists exclusively of first-order spatio-temporal image derivatives. The estimator consists in a sequence of several bidimensional regression problems, easily solved by known statistical tools. The use of robust statistic tools in a tractable manner, and the ability to cope with general motion are other salient aspects of this work.

2. Egomotion Estimation

Consider a camera with linear velocity $\mathbf{t} = [U\ V\ W]^T$, angular velocity $\boldsymbol{\omega} = [\omega_1\ \omega_2\ \omega_3]^T$ and the scene depth denoted by Z. The motion field induced in the image plane can be calculated at every image site $(x,\ y)$. Assuming that the scene is composed of textured objects under slowly varying illumination, the brightness-constancy hypothesis [3] holds. Thus we can only estimate the component of the optical flow, $\mathbf{v}$, along the image gradient direction, $\mathbf{n}$. This projection, commonly designated by the *normal-flow*, v_n, depends on the motion parameters as follows:

$$\begin{aligned} v_n(\mathbf{x}) &= \rho(\mathbf{x})\mathbf{n}(\mathbf{x})\cdot(x-\sigma,\ y-\eta) + \mathbf{n}^{\mathbf{t}}(\mathbf{x})\mathbf{B}(\mathbf{x})\boldsymbol{\omega} \\ \mathbf{B}(\mathbf{x}) &= \begin{bmatrix} \frac{xy}{f} & -(\frac{x^2}{f}+f) & y \\ (\frac{y^2}{f}+f) & -\frac{xy}{f} & -x \end{bmatrix} \end{aligned} \tag{1}$$

where $\rho = W/Z$ and $(\sigma,\ \eta) = (fU/W,\ fV/W)$ is the Focus of Expansion (FOE).

To define an egomotion estimator using directly the normal flow, we consider some of its properties. Consider a

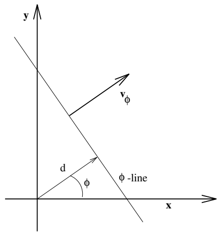


Figure 1. Each v_ϕ lies on a ϕ-line, defined by a vector with an orientation ϕ and length d.

generic image straight line, ϕ-line, defined by a vector with orientation $\phi \in \left]-\pi/2, \pi/2\right]$ and length d (Figure 1). The normal flow vectors $\mathbf{v}_\phi$, perpendicular to this line, depend on the motion and line parameters, as follows:

$$\begin{aligned} v_\phi = &\underbrace{\rho(d-(\sigma \mathrm{c}\phi + \eta \mathrm{s}\phi))}_{\textit{translational component}} + \\ &\underbrace{+\left(\frac{d}{f}(\omega_1 \mathrm{c}\phi + \omega_2 \mathrm{s}\phi) + \omega_3\right) r + \left(\frac{d^2}{f}+f\right)(\omega_1 \mathrm{s}\phi - \omega_2 \mathrm{c}\phi)}_{\textit{rotational component}} \end{aligned} \tag{2}$$

*This work was partially funded by the PRAXIS/3/3.1/TPR/23/94, JNICT-PBIC/TPR/2550/95 and EC ESPRIT-LTR NARVAL projects and a grant from the PRAXIS-XXI research program.

where $r = -x \sin\phi + y \cos\phi$, $(\mathrm{c}\phi,\ \mathrm{s}\phi) = (\cos\phi,\ \sin\phi)$, and $v_\phi = \mathbf{v}(\mathbf{x}) \cdot (\cos\phi, \sin\phi)$[1]. The rotational component of the normal flow vectors $\mathbf{v}_\phi$ is affine in r with the constant term defining a linear combination of ω_1 and ω_2. The translational component vanishes when the ϕ-line goes through the FOE, i.e., the normal flow vectors perpendicular to a line joining **x** and the FOE are *solely* dependent on the rotation.

As each normal flow vector defines one ϕ-line, we work on the **lines topological space $\boldsymbol{\mathcal{L}} \equiv \{\mathbf{d}, \boldsymbol{\phi}\}$** spanning the universe of all possible ϕ-lines. The $\mathcal{L}$-space is bounded and can be discretized according to the image grid. Each line $\{d, \phi\}$ accumulates contributions of a finite set of normal flow vectors $\mathbf{v}_\phi$. Additionally, each vector is linked to a coordinate, r, on the associated line[2].

Next, we present a low complexity estimator based on the simple regression problem, by applying sequentially bidimensional estimators, defined on the $\mathcal{L}$-space.

The normal flow observations $\mathbf{v}_\phi$ collected along a ϕ-line $\{d_i, \phi_j\}$, depend on the desired motion parameters, according to the equation (2) that can be written as follows:

$$v_{\phi_{ij}} = a_{ij} r + b_{ij} + e \tag{3}$$

where a_{ij} and b_{ij} are the affine parameters of the rotational component of the vectors $v_{\phi_{ij}}$, and e is an associated error (including the translational component and noise)[3].

The problem is to estimate both a_{ij} and b_{ij}. Various estimators have been proposed to detect and remove outliers [5]. We adopted the Least Median of Squares (LMS) estimator, $\min_{\hat{a}_{ij}\hat{b}_{ij}} \mathrm{med}_k R_k^2$, where the residual $R = v_{\phi_{ij}} - (a_{ij} r + b_{ij})$ and k indexes the observations on line $\{d_i, \phi_j\}$. In this bidimensional regression problem, the LMS estimation can be solved by a non-iterative algorithm, as presented in [5].

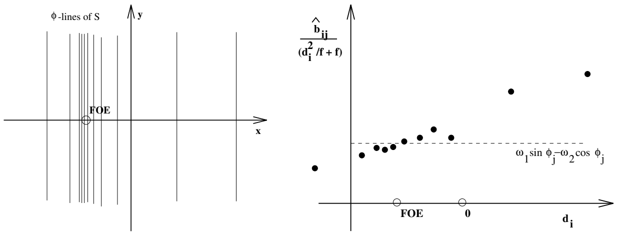
When determining an LMS estimate of $\hat{a}_{ij}$ and $\hat{b}_{ij}$, the residuals, R_k, are compared to an estimate, ϵ_{ij}, of the error scale, given by $\epsilon_{ij} = \sqrt{\mathrm{med}_k R_k^2}$. The value of ϵ_{ij} is related to the robust standard deviation of the residuals, and is large either when (i) the translational component is dominant and strongly non-linear or (ii) the observations are corrupted by high variance noise.

The translation influence on ϵ_{ij} vanishes when the ϕ-line intersects the FOE, because the non-linear behavior of translational component is minimized. In some cases, the translational component may depend linearly in r. Then, although ϵ_{ij} can be small, the estimates $\hat{a}_{ij}$, $\hat{b}_{ij}$ will depend on the translational motion. This undesirable effect can be addressed later when the rotation parameters are estimated.

First we eliminate the candidate ϕ-lines with large ϵ_{ij}. We use a simple criterium to select the set S of the valid ϕ-lines: $S = \{\forall \{d_i, \phi_j\} : \epsilon_{ij} < \epsilon^*\}$, where $\epsilon^* = \mathrm{med}_{ij}\epsilon_{ij}$. To recover the motion parameters, we use uniquely the values $\hat{b}_{ij}$ of ϕ-lines $\in S$. According to equation (2), we have

$$\hat{b}_{ij} = \left(\frac{d_i^2}{f} + f\right)(\omega_1 \sin\phi_j - \omega_2 \cos\phi_j) + \delta_{ij}, \tag{4}$$

where δ_{ij} is the robust mean of the error e of equation (3). In Figure 2a, we consider a subset of S with the same



(a) (b)

Figure 2. (a) Family of ϕ-lines included in S with the same orientation. (b) Evolution of $\hat{b}_{ij}$, scaled by $\left(\frac{d_i^2}{f} + f\right)$.

orientation ϕ. Due to the ϵ^*-selection applied before, the resulting ϕ-lines are concentrated in a neighborhood of the FOE. Figure 2b illustrates the evolution of $\hat{b}_{ij}$ for different d_i and fixed ϕ_j. When the ϕ-lines are close to the FOE, δ_{ij} approaches zero; otherwise, δ_{ij} spreads around some non-zero value. Figure 2b also shows that δ_{ij} has opposite signs on the left and on the right sides of the line crossing the FOE. This particular result is due to the known translational component behavior around the FOE, assuming that depth is always positive.

Next, using the estimated information from ϕ-lines $\in S$, we minimize two separate cost functions which depend on the translation and the rotation, as follows:

$$\min_{\hat{\sigma}\hat{\eta}} \mathrm{med}_{\phi \in S} \left(d_i - (\sigma \cos\phi_j + \eta \sin\phi_j)\right)^2 \tag{5}$$

$$\min_{\hat{\omega_1}\hat{\omega_2}} \mathrm{med}_{\phi \in S} \left(\hat{b}_{ij} - \left(\frac{d_i^2}{f} + f\right)(\omega_1 \sin\phi_j - \omega_2 \cos\phi_j)\right)^2 \tag{6}$$

By selecting the ϕ-lines with smallest residuals both for the first and the second estimators, we define a subset of S under the constraint that (i) the ϕ-lines are close to a good candidate of the FOE, and (ii) the corresponding normal flow vectors, $\mathbf{v}_{\phi_{ij}}$, are linearly dependent on some rotation vector.

Let R_{ij}^T and R_{ij}^ω be the residuals computed with $(\hat{\sigma}, \hat{\eta})$ and $(\hat{\omega_1}, \hat{\omega_2})$, respectively. Defining $\epsilon^T = \sqrt{\mathrm{med}_{\phi \in S} {R_{ij}^T}^2}$

[1] Notice that $v_\phi = \pm v_n$, where v_n is the normal flow. The two values have opposite sign, whenever the image gradient angle corresponds to the suplementar angle of ϕ.

[2] The point $r = 0$ is the nearest line point to the image plane origin.

[3] Notice that e may have a non-zero mean and/or a non-gaussian distribution.

and $\epsilon^{\omega} = \sqrt{\text{med}_{\phi \in S} {R^{\omega}_{ij}}^2}$ we have a new subset S' given by:

$$S' = \left\{ \forall \phi_{ij} \in S : \left(\frac{R^T_{ij}}{\epsilon^T} \right)^2 + \left(\frac{R^{\omega}_{ij}}{\epsilon^{\omega}} \right)^2 < 1 \right\} \qquad (7)$$

To improve the quality of $(\hat{\sigma}, \hat{\eta}, \hat{\omega_1}, \hat{\omega_2})$, we apply again the estimators (5) and (6) to a selection of vectors, $\mathbf{v}_{\phi_{ij}}$, lying on the ϕ-lines $\in S'$. By using the equation (2), the remaining parameter ω_3 can be easily computed since the other motion parameters are already known.

3. Results

We have conducted a series of experiments of different difficulty, using real image data. Five frames are used to compute the normal flow field. The $\mathcal{L}$-space is a discrete domain, bounded by the image dimensions, discretized every 0.2 radians along the ϕ-axis and every 5 pixels along the d-axis. We used the image top-left corner as the $\mathcal{L}$-space origin, as it can be defined arbitrarily on the image plane[4].

In the figures below, we represent each ϕ_{ij}-line by a single point, $(d_i \cos \phi_j,\ d_i \sin \phi_j)$. This simple representation provides a clear illustration of the estimation process.

Sequence 1, pure translation: In this sequence, the camera undergoes a pure translational motion in a real cluttered scene. Figure 3 shows uniquely the first selection step of ϕ-lines. Since we minimize the cost function in (5), the corresponding dots given by $(d_i \cos \phi_j,\ d_i \sin \phi_j)$ tend to a circular shape having the $\mathcal{L}$-space origin and the FOE as diametrical opposite points. Table 1 shows the final rotation and translation estimates, in image coordinates. As expected, the estimated rotation is small. The angular error between the real translation vector and its estimate is also small ($<$ 5 degrees).

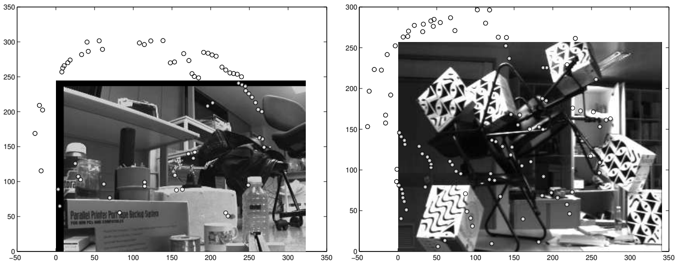


Figure 3. Resulting set S from Sequences 1 (left) and 2 (right). The $\mathcal{L}$-space origin is located at the image top-left corner.

[4]Notice that in the numerical computation of the rotational estimators we must take into account the actual origin (the intersection of the optical axis in the image plane).

Motion	Estimates	True values
Seq.1: $(\boldsymbol{\omega_1},\ \boldsymbol{\omega_2})$	$(0.0009,\ 0.0005)$	$(0.0,\ 0.0)$
Seq.1: $(\boldsymbol{\sigma},\ \boldsymbol{\eta})$	$(226.6,\ 80.5)$	$(245,\ 65)$
Seq.2: $(\boldsymbol{\omega_1},\ \boldsymbol{\omega_2})$	$(0.0000,\ -0.0051)$	$(0.0,\ -0.005)$
Seq.2: $(\boldsymbol{\sigma},\ \boldsymbol{\eta})$	$(140,\ 120)$	$(146,\ 130)$

Table 1. Estimation results. Rotational motion in rad/frame and FOE in pixels.

Sequence 2, rotation with translation: In this sequence, the camera undergoes a translational motion and rotates around the Y axis, with $\omega_2 = -0.005$ rad/frame. The rotational component of the normal flow field is dominant in relation to the translational one. Therefore, the ϕ-lines $\in S$ (Figure 3) are more spread than usual, and the rotation parameters estimates are very accurate (Table 1). After completing the estimation process, the FOE estimate is close to the real value.

4. Conclusions

We addressed the problem of egomotion estimation for a monocular observer with arbitrary translation and rotation. Our approach is uniquely based on spatio-temporal image derivatives. The egomotion process is based on the subdivision of the problem in various bidimensional regression problems. An LMS estimator of low complexity is applied in order to estimate the motion parameters and improve the estimation robustness.

Robustness is achieved in two ways. First, we use exclusively the normal flow field together with robust statistics methods. Secondly, the problem is defined in a topological space — the $\mathcal{L}$-space — such that the estimates are supported by a large set of image data, as opposed to some previous approaches.

References

[1] C. Fermuller. Passive navigation as a pattern recognition problem. *IJCV*, 14(2):147–158, March 1995.

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

[3] B. Horn and B. Shunck. Determining optical flow. *Art. Intelligence*, 17:185–203, 1981.

[4] F. L. O. Faugeras and G. Toscani. Motion and structure from motion from point and line matches. In *Proc. 1st Intern. Conf. Comput. Vision*, London, June 1987.

[5] P. Rousseeuw and A. Leroy. *Robust Regression & Outlier Detection*. John Wiley & Sons, Inc, 1987.

[6] C. Silva and J. Santos-Victor. Robust egomotion estimation from the normal flow using search subspaces. *IEEE Transactions on PAMI*, 19(9):1026–1034, September 1997.