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. Extensive experimental tests are presented and discussed.

Keywords: Normal flow, egomotion, robust regression methods
Track: Computer Vision.

1 Introduction

The interpretation of visual motion has attracted a good deal of attention of computer vision researchers, for a long time. In particular it conveys information about the scene structure and the observer egomotion.

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 [7] or by computing the dense optical flow field [4].

Alternatively, the direct methods rely exclusively on the normal flow field, or first order spatio temporal image derivatives [6, 2, 3], thus overcoming the ill-posedness nature of the flow estimation image matching problems [1].

We extend previous work [9, 10], 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 Problem Statement

Consider a camera with linear velocity $\mathbf{t} = [U\ V\ W]^T$, angular velocity $\boldsymbol{\omega} = [\omega_1\ \omega_2\ \omega_3]^T$ and that the scene depth at each point is denoted by Z. The motion field induced in the image plane, can be calculated at every pixel by the following well-known vector equation [4] :

$$\mathbf{v}(\mathbf{x}) = \rho(\mathbf{x})(\mathbf{x} - \boldsymbol{\Sigma}) + \mathbf{B}(\mathbf{x})\boldsymbol{\omega}, \quad \mathbf{B}(\mathbf{x}) = \begin{bmatrix} \frac{xy}{f} & -(\frac{x^2}{f} + f) & y \\ (\frac{y^2}{f} + f) & -\frac{xy}{f} & -x \end{bmatrix} \tag{1}$$

where $\mathbf{x} = [x\ y]^T$ denotes an image site, $\mathbf{v}(\mathbf{x}) = [u(\mathbf{x})\ v(\mathbf{x})]^T$ is the optical flow along the x and y axis. The Focus of Expansion (FOE), $\boldsymbol{\Sigma} = [\sigma\ \eta]^T = [fU/W, fV/W]^T$ is the projection of the observer linear velocity, $\mathbf{t}$, on the image plane. The function $\rho(\mathbf{x})$ is the inverse of the time-to-crash, given by $\rho(x, y) = W/Z(x, y)$.

Assuming that the scene is composed of textured objects under slowly varying illumination, the familiar brightness-constancy hypothesis [5] holds. This implies that we can only estimate the component of the optical flow, $\mathbf{v}$, along the image gradient direction, vector $\mathbf{n}$. This projection $\mathbf{n} \cdot \mathbf{v}$ is commonly designated by the *normal-flow*, v_n, and depends on the motion parameters as follows:

$$v_n(\mathbf{x}) = \rho(\mathbf{x})\mathbf{n}(\mathbf{x}) \cdot (\mathbf{x} - \boldsymbol{\Sigma}) + \mathbf{n}^{\mathbf{t}}(\mathbf{x})\mathbf{B}(\mathbf{x})\boldsymbol{\omega} \tag{2}$$

To define an egomotion estimator using directly the normal flow, we consider some of its properties. First of all, equation (2) shows that the normal flow vectors perpendicular to a line joining $\mathbf{x}$ and the FOE, are *solely* dependent on the camera rotation. Hence, estimating the rotation (collecting data on these image sites) consists on a linear problem, by minimizing the sum of the squared residuals.

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

$$v_\phi = \underbrace{\rho(d - (\sigma \cos\phi + \eta \sin\phi))}_{\textit{translational component}} + \underbrace{\left(\frac{d}{f}(\omega_1 \cos\phi + \omega_2 \sin\phi) + \omega_3\right) r + \left(\frac{d^2}{f} + f\right)(\omega_1 \sin\phi - \omega_2 \cos\phi)}_{\textit{rotational component}}, \tag{3}$$

where $r = -x \sin\phi + y \cos\phi$, and $v_\phi = \mathbf{v}(\mathbf{x}) \cdot (\cos\phi, \sin\phi)$[1]. The rotational component of the

[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.

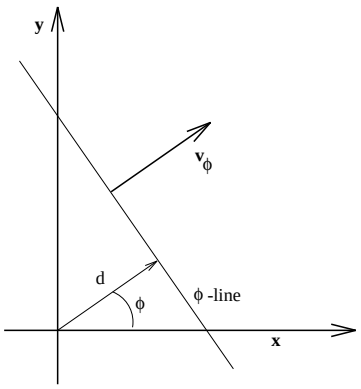

Figure 1: Each v_ϕ lies on a ϕ-line, defined by a vector with an orientation ϕ and a length d.

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.

In previous work [9, 10], we presented egomotion estimators based on search subspaces, minimizing the residual due to the translational component (usually non-linear). We now propose a more general and global method to detect the motion parameters.

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. Similarly, 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]. We propose an egomotion estimator based on the $\mathcal{L}$-space, thus allowing the use of nearly all the available image data.

3 Egomotion Estimation

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

Very often in computer vision there is the need to use estimation algorithms to recover several parameters. Most researchers apply least square (LS) techniques over the residual of some cost function. However, LS methods are optimal only when the observation error is zero-mean with normal distribution. This is rarely the case in many problems, and a few data outliers can ruin the quality of the estimates.

We will solve a simple regression problem on the normal flow observations $\mathbf{v}_\phi$ collected along

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

a ϕ-line $\{d_i, \phi_j\}$. The flow on the ϕ-line depends on the desired motion parameters, according to the general equation (3) that can be written as follows:

$$v_{\phi_{ij}} = a_{ij} r + b_{ij} + e \tag{4}$$

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 (see [8] for an extended introduction to robust estimation). We adopted the Least Median of Squares (LMS) estimator:

$$\min_{\hat{a}_{ij}\hat{b}_{ij}} \text{med}_k R_k^2 \tag{5}$$

where the residual $R = v_{\phi_{ij}} - (a_{ij} r + b_{ij})$ and k indexes the observations in line $\{d_i, \phi_j\}$. In this bidimensional regression problem, the LMS estimation can be solved by a non-iterative algorithm, as presented in [8].

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. The value of ϵ_{ij} must depend only on the "good" data, and can be computed robustly with the residuals R_k: $\epsilon_{ij} = \sqrt{\text{med}_k R_k^2}$. An observation is considered an outlier if the corresponding residual, R_k, is larger than ϵ_{ij}. The value of ϵ_{ij} is related to the robust standard deviation of the residuals, and is larger 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[4]. This undesirable effect can be addressed later when the rotation parameters are estimated.

The first step of the algorithm eliminates candidate ϕ-lines with large ϵ_{ij}. We use a simple criterium to select the set S of the valid ϕ-lines:

$$S = \{\forall \{d_i, \phi_j\} : \epsilon_{ij} < \epsilon^*\}, \qquad \epsilon^* = \text{med}_{ij} \epsilon_{ij}.$$

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

[4]e.g the estimated value $\hat{b}_{ij}$, rather than being a pure linear combination of two rotational parameters, will be contaminated by structure and translation.

To recover the motion parameters, we use uniquely the values $\hat{b}_{ij}$ of ϕ-lines $\in S$. According to equation (3), we have

$$\hat{b}_{ij} = \left(\frac{d_{ij}^2}{f} + f\right)(\omega_1 \sin\phi_{ij} - \omega_2 \cos\phi_{ij}) + \delta_{ij}, \tag{6}$$

where δ_{ij} is the robust mean of the error e of equation (4).

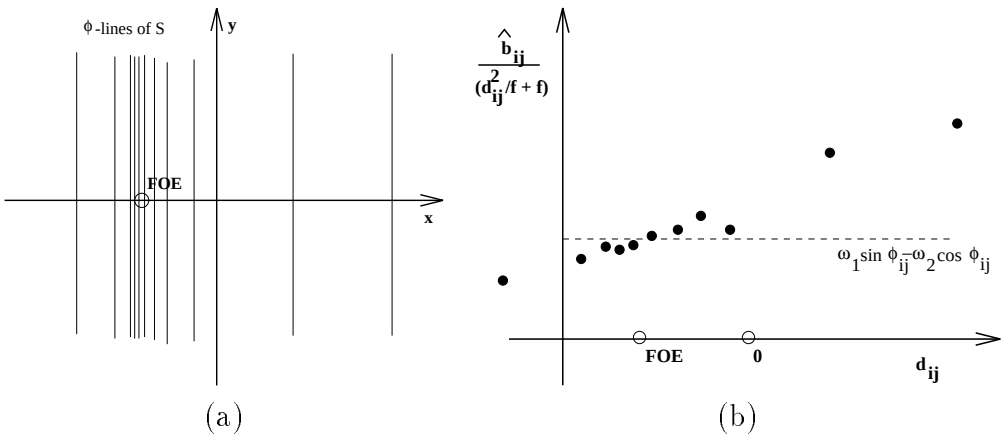

Figure 2: (a) Family of ϕ-lines included in S with the same orientation. (b) The corresponding evolution of the b_{ij} estimates, scaled by $\left(\frac{d_{ij}^2}{f} + f\right)$.

In Figure 2a, we consider a subset of S with the same 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_{ij} and fixed ϕ_{ij}. 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 can be easily proved algebraically by using equation (2) and is due to the known translational component behavior around the FOE, assuming that depth is always positive. This property can be used to identify lines going through the FOE, by a geometric search algorithm[5]. In this paper, we incorporate naturally these features into our estimation approach, assuming that the scene is sufficiently rich in terms of texture and depth variability.

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:

$$\min_{\hat{\sigma}\hat{\eta}} \text{med}_{\phi \in S}(d_{ij} - (\sigma \cos\phi_{ij} + \eta \sin\phi_{ij}))^2 \tag{7}$$

[5]This geometric approach may be specially suitable when there are a few observations in the ϕ-lines that go through the FOE (or close), due to poorly textured images.

$$\min_{\hat{\omega_1}\hat{\omega_2}} \text{med}_{\phi \in S} \left(\hat{b}_{ij} - \left(\frac{d_{ij}^2}{f} + f \right) (\omega_1 \sin \phi_{ij} - \omega_2 \cos \phi_{ij}) \right)^2 \tag{8}$$

The first estimator determines the FOE, $(\hat{\sigma}, \hat{\eta})$, by robustly minimizing the translation component of $\mathbf{v}_{\phi_{\mathbf{ij}}}$ vectors (or, equivalently, δ_{ij} term). The second estimator recovers the rotation by using the estimates $\hat{b}_{ij}$.

When estimating the motion parameters through equations (7, 8) we also compute the corresponding residuals. 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, $\mathbf{v}_{\phi_{\mathbf{ij}}}$, 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 $\epsilon^T = \sqrt{\text{med}_{\phi \in S} {R_{ij}^T}^2}$ and $\epsilon^\omega = \sqrt{\text{med}_{\phi \in S} {R_{ij}^\omega}^2}$ we have a new subset S' given by:

$$S' = \left\{ \forall \phi_{ij} \in S : \left(\frac{R_{ij}^T}{\epsilon^T} \right)^2 + \left(\frac{R_{ij}^\omega}{\epsilon^\omega} \right)^2 < 1 \right\} \tag{9}$$

To improve the quality of $(\hat{\sigma}, \hat{\eta}, \hat{\omega_1}, \hat{\omega_2})$, we apply again the estimators (7) and (8) to a selection of vectors, $\mathbf{v}_{\phi_{\mathbf{ij}}}$, lying on the ϕ-lines $\in S'$. The remaining parameter ω_3 can be easily computed if the other motion parameters are already known. The following special cases deserve additional attention:

Pure rotation: In this case, all normal flow vectors depend linearly on the rotation parameters. Thus, the ϕ-lines of S (after ϵ^*-selection) are widely spread (more or less uniformly). Hence, estimating the FOE using estimator (7) leads to a large standard deviation described by ϵ^T. To detect the pure-rotation case, one can establish two decision thresholds for ϵ^T and ϵ^ω. If ϵ^T value is above some threshold and ϵ^ω is below another one, then we decide that the camera motion is a pure rotation.

The FOE at the infinity: In the approach proposed before, the FOE location can lie outside the image domain. However, when the translation velocity is almost parallel to the image plane, the FOE goes to infinity. In this case, one can determine the direction of the vector (σ, η) rather than the exact FOE location. According to the $\mathcal{L}$-space, if the FOE is at infinity, then the ϕ-lines that intersect the FOE are parallel and the direction ϕ is orthogonal to (σ, η).

Assume that the regression data matrix, $\mathcal{D}$, is the collection of the vectors $(\cos\phi_{ij}, \sin\phi_{ij})$, for all ϕ-lines $\in S$. If the FOE is unbounded, $\mathcal{D}$ is approximately singular, due to an existing dominant direction on the ϕ-lines. Therefore, before applying directly the estimator (7), we compute the two singular values of $\mathcal{D}$. If one of them is very small relative to the other, we consider that the FOE is approximately at the infinity and its direction given by the dominant singular vector. Remember that this result must be validated with a low variance of the rotational estimator (8), in order to guarantee robustness.

4 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[6].

In the figures below, we represent each ϕ_{ij}-line by a single point, $(d_{ij}\cos\phi_{ij},\ d_{ij}\sin\phi_{ij})$. 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 the two selection steps of ϕ-lines. Since we minimize the cost function in (7), the corresponding dots given by $(d_{ij}\cos\phi_{ij},\ d_{ij}\sin\phi_{ij})$ tend to a circular shape having the $\mathcal{L}$-space origin and the FOE as diametrical opposite points. Table 1 shows the rotation and translation estimates, in image coordinates. As expected, the

Motion	**Estimates**	**True values**
$(\boldsymbol{\omega_1},\ \boldsymbol{\omega_2})$	(0.00092, 0.00051) rad/fr	(0.0, 0.0) rad/fr
$(\boldsymbol{\sigma},\ \boldsymbol{\eta})$	(226.6, 80.5) pix	(245, 65) pix

Table 1: Estimation results associated to the first sequence.

estimated rotation is small. The angular error between the real translation vector and its estimate is also small (< 5 degrees).

Sequence 2, pure translation: This experiment uses the "coke"-sequence (Figure 4). Few information is available about the camera calibration, like the focal length and the principal

[6] 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).

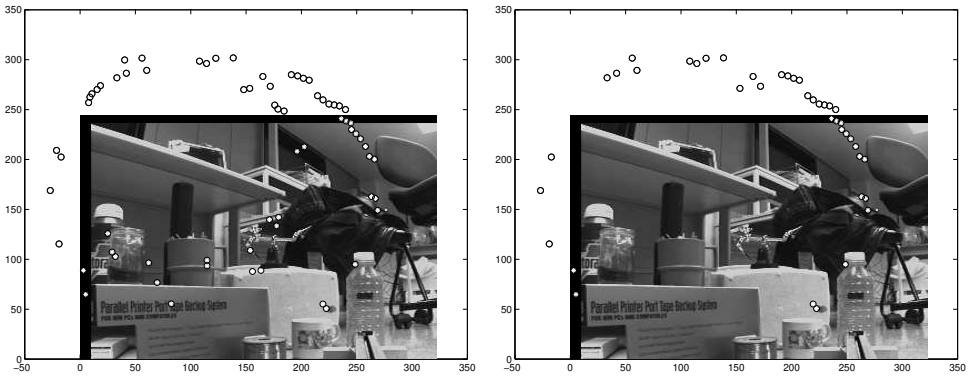

Figure 3: Sequence 1: Left - set S. Right - resulting set S'. The $\mathcal{L}$-space origin is located at the image top-left corner.

point. An approximated location of the FOE was obtained visually for comparison purposes. This trial shows that the algorithm can cope well with the case of pure translation. Table

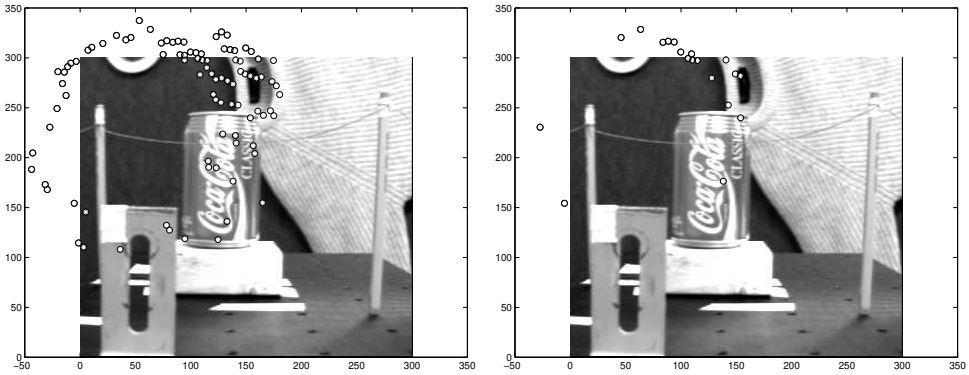

Figure 4: Sequence 2: Left - set S. Right - resulting set S'. The $\mathcal{L}$-space origin is located at the image top-left corner.

Motion	**Estimates**	**Empirical values**
$(\boldsymbol{\omega_1},\ \boldsymbol{\omega_2})$	$(-0.00001,\ -0.00007)$ rad/fr	$(0.0,\ 0.0)$ rad/fr
$(\boldsymbol{\sigma},\ \boldsymbol{\eta})$	$(124.4,\ 160.0)$ pix	$(130,\ 175)$ pix

Table 2: Estimation results associated to the second sequence.

2 shows that the rotation estimates are very small and the estimated FOE is close to the corresponding empirical value.

Sequence 3, 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 5 on the left) are more spread than usual, and the rotation parame-

ters estimates are very accurate (Table 3). After completing the estimation process, the FOE

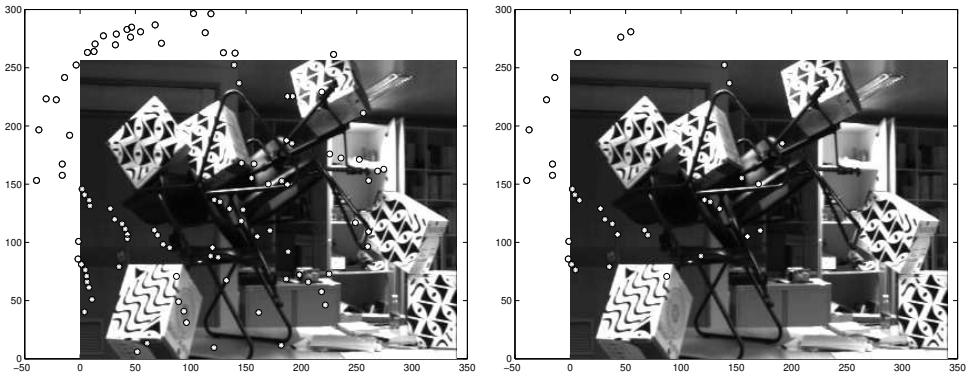

Figure 5: Sequence 3: Left - set S. Right - resulting set S'. The $\mathcal{L}$-space origin is located at the image top-left corner.

estimate is close to the real value.

Motion	**Estimates**	**Real values**
$(\boldsymbol{\omega_1},\ \boldsymbol{\omega_2})$	$(0.0000,\ -0.0051)$ rad/fr	$(0.0,\ -0.005)$ rad/fr
$(\boldsymbol{\sigma},\ \boldsymbol{\eta})$	$(140,\ 120)$ pix	$(146,\ 130)$ pix

Table 3: Estimation results associated to the third sequence.

Sequence 4, FOE at the infinity: The last experiment with the "trees"-sequence, consists on a pure translation parallel to the image plane. For this special case, the ϕ-line dots

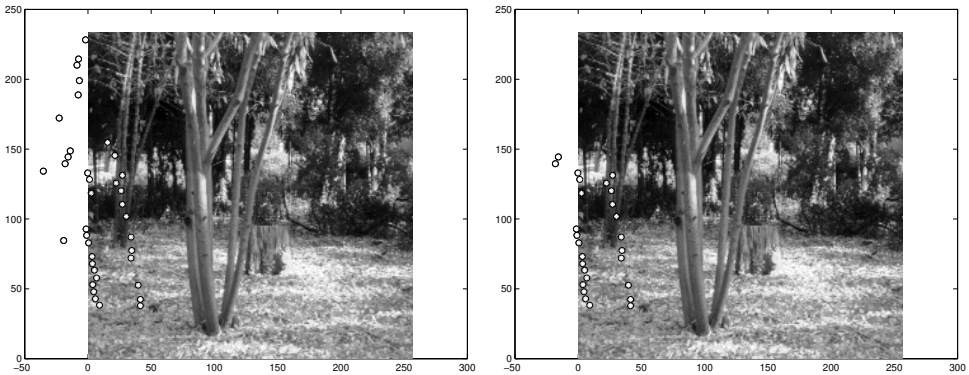

Figure 6: Sequence 4: Left - set S. Right - resulting set S'. The $\mathcal{L}$-space origin is located at the image top-left corner.

(Figure 6) form a straight line (degenerate case) perpendicular to FOE direction. Therefore, the Figure 6 shows that the FOE direction is close to the horizontal, as expected. The estimated rotation values were negligible ($\approx 10^{-4}$).

5 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.

The robustness of the approach is illustrated with a series of experiments using real images.

References

[1] Y. Aloimonos, I. Weiss, and A. Banddophaday. Active vision. *Int. Journal of Computer Vision*, 1(4):333–356, January 1988.

[2] C. Fermuller. Passive navigation as a pattern recognition problem. *Int. Journal of Computer Vision*, 14(2):147–158, March 1995.

[3] C. Fermuller and Y. Aloimonos. The role of fixation in visual motion analysis. *Int. Journal of Computer Vision*, 11(2):165–186, October 1993.

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

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

[6] B.K.P. Horn and E.J. Weldon. Direct methods for recovering motion. *International Journal of computer Vision*, 2(1):51–76, 1988.

[7] F. Lustman 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.

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

[9] C. Silva and J. Santos-Victor. Direct egomotion estimation. In *Proc. of the 13th Int. Conference on Pattern Recognition*, Vienna,Austria, August 1996.

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