# 3-D TRANSLATIONAL MOTION ESTIMATION FROM 2-D DISPLACEMENTS

*C. Garcia and G. Tziritas*

Department of Computer Science, University of Crete
P.O. Box 2208, 71409 Heraklion, Greece
E-mails: {cgarcia,tziritas}@csd.uoc.gr

## ABSTRACT

Recovering 3-D motion parameters from 2-D displacements is a difficult task, given the influence of noise contained in these data, which correspond at best to a crude approximation of the real motion field. The need for stability of the system of equations to solve is therefore essential. In this paper, we present a novel method based on an unbiased estimator that aims at enhancing this stability and strongly reduces the influence of noise contamination. Experimental results using synthetic and real optical flows are presented to demonstrate the effectiveness of our method in comparison to a set of selected methods.

## 1. INTRODUCTION

The estimation of 3-D motion parameters from a sequence of images is a fundamental task in image analysis with numerous applications, such as egomotion and time-to-contact estimation for mobile robots, video segmentation, depth layering, or video content description. Most methods for 3-D motion analysis begin by extracting two-dimensional motion information. Many algorithms have been proposed for extracting 3-D motion parameters from optical flow. A detailed review is proposed by Heeger and Jepson in [1]. Heeger and Jepson [1] minimize a residual function where depth and rotation parameters are eliminated in order to obtain a measure of error as a function of translation which is then analyzed to select the correct translation. Lobo and Tsotsos [2] propose a voting scheme based on triplets of points using the Collinear Point Constraint for cancelling rotation and finding the focus of expansion. Daniilidis [3] makes use of fixation on a scene point and projection of the spherical motion field on two latitudinal directions to decouple the motion parameter space, searching then along meridians of the image sphere.

One main problem in correctly estimating the camera motion parameters is the fact that the 2-D motion field usually contains noisy data and outliers, making most of the above mentioned methods unstable. The set of incorrect data can be even larger if independent motions exist throughout the image sequence. The negative effects of this set of outliers on motion estimation increase with the complexity of the motion model which is used to describe the camera motion. Komodakis and Tziritas [4] proposed a robust estimation method to cope with the set of outliers and the use of a hierarchy of motion models sequentially tested. In this paper, we focus on improving the motion parameter estimation in the case of translational motion. Section 2 describes the equations linking the projected 2-D motions and 3-D motions inside the image sequence, which yields an overdetermined system of linear equations in the translation case. Section 3 presents our approach, based on the projection of the equations' coefficients into a different space, chosen appropriately in order to reduce the influence of noise contamination. In Section 4, experimental results using synthetic noisy optical flows are analyzed in order to compare the selected methods and to show the superiority of our approach. Experimental results using real optical flow are also presented. Finally, conclusions are drawn.

## 2. 3-D MOTION PARAMETERS FROM 2-D DISPLACEMENTS

### 2.1. Optical flow

In this paper we only consider the pure 3-D translational motion. The 2-D motion vector $(u, v)$ at an image point $(x, y)$ can be expressed using the instantaneous 3-D translation vector $(T_X, T_Y, T_Z)$

$$u = \frac{-T_x f + xT_z}{Z} \quad \text{and} \quad v = \frac{-T_y f + yT_z}{Z} \qquad (1)$$

By eliminating $Z$ from the motion field equations (1) and introducing, in the case of $T_z \neq 0$, the notation $(\alpha = T_x f/T_z, \beta = T_y f/T_z)$, we obtain for all points $i$:

$$v(i)\alpha - u(i)\beta = x(i)v(i) - y(i)u(i) \qquad (2)$$

The point $(\alpha, \beta)$ is called *Focus of Expansion* (FOE) and corresponds to the point of intersection of the lines supporting the motion vectors defined by the translational components. This case of 3-D translation will be referred as to *full translation*.

In the case of $T_z = 0$, the FOE is at infinity. Only the direction of translation may be recovered. This direction is defined by the ratio $\gamma = T_y/T_x$ (or $T_x/T_y$). This case of 3-D motion will be referred as to *panning*.

In both cases, the parameter estimation consists in solving the corresponding overdetermined system, where all coefficients are noisy, given that they depend on $u$ and $v$. Indeed, the observed optical flow field is a very crude approximation of the motion field, whatever method for computing it is used. An interesting review of optical flow techniques including performance analysis is presented in [5].

0-7803-6725-1/01/$10.00 ©2001 IEEE

### 2.2. Point correspondences

Let us now consider the discrete case where point correspondences have been obtained. Let $(x', y')$ be at time $t'$ the point corresponding to $(x, y)$ at time $t$. Let us denote again by $(T_X, T_Y, T_Z)$ the 3-D translational displacement. We then obtain the relations of image point coordinates

$$x' = \frac{xZ - fT_X}{Z - T_Z}, \quad y' = \frac{yZ - fT_Y}{Z - T_Z}. \qquad (3)$$

By eliminating $Z$ from the above two correspondence equations, if $T_Z \neq 0$, we obtain one linear equation for each point correspondence, which is quite similar to the equation obtained with the optical flow vector,

$$-(y'-y)\alpha+(x'-x)\beta = x'y-xy' = (x'-x)y-(y'-y)x. \quad (4)$$

From the algebraic point of view, if we denote $u = x' - x$ and $v = y' - y$, we have exactly the same equations.

### 2.3. Existing methods for solving overdetermined systems

Several main techniques have been proposed for solving overdetermined linear systems. The simplest and therefore most often used error minimizing technique is Least Squares (LS). Although it offers a simple technique for solving the problem, the provided estimate is biased.

In the case of errors affecting the equation coefficients the *Total Least Square* (TLS) algorithm aims at finding the best solution of the overdetermined system of equations. This is performed via classical eigenanalysis on Singular Value Decomposition. The smallest eigenvalue is selected and the solution depends on the corresponding eigenvector. In the case where there are multiple small eigenvalues, instability appears in the solution of TLS.

Least-squares-based estimators may be completely perturbed by a few outliers [6]. The goal of positive-breakdown methods is robustness against the presence of several unannounced outliers that may have occurred anywhere in the data. There are several types of high-breakdown robust methods, in particular the Least Median of Squares (LMedS) and the M-estimators. An interesting review is given in [7].

The M-estimators method [8] can be reduced to a reweighted least-squares (RLS) technique. It is used for 3-D motion estimation in [4]. In our approach, among different M-estimators we selected the Tukey estimator. The Tukey's weighting function uses a scale parameter $c$, chosen in our implementations as a function of the median of the residuals.

## 3. THE OPTIMAL PROJECTION METHOD

All the previous methods, *i.e.*, LS, TLS, RLS, try to solve for the motion parameters using a large set of equations where the coefficients are very unstable, given the noise affecting the optical flow vectors $u$ and $v$. This is the key observation that has given rise to our method. We search for equations whose coefficients are optimal according to criteria derived from the supposed noise model of the optical flow. They are basically obtained by first projecting the vector of coefficients of each original equations into a space with a chosen basis of vectors. This scheme greatly reduces the influence of noise contamination in the new set of equation coefficients. Moreover, unlike all the above mentioned methods, our estimator is designed to be unbiased.

### 3.1. Noise in optical flow observations

The proposed method is based on the model of the noise affecting the optical flow data. We suppose that the two components of the motion field $u$ and $v$ are perturbed by additive zero-mean Gaussian noise. The two noise processes are assumed to be independent, and each of them is assumed to be spatially uncorrelated. This last property is not necessary for obtaining an unbiased estimator, but it is included for simplifying the variance expressions.

The variance of the noise is supposed to be either constant or proportional to the square of the corresponding component. This last model seems compatible with the probability distribution of optical flow proposed in [9] and the observations made in the review of optical flow techniques by Barron et al. [5]. Similar noise models are used in [10, 2, 11]. Considering the proposed noise model, we have:

$$u(i) = \mu(i) + N_1(i) \quad \text{and} \quad v(i) = \nu(i) + N_2(i) \qquad (5)$$

where $i$ indexes the image points where an optical flow vector is defined and $\mu(i)$ and $\nu(i)$ are the ideal optical flow components at point $i$. When the "proportional" model is used the noise processes $N_1$ and $N_2$ are such that:

$$E\left\{N_1^2(i)\right\} = \sigma^2\mu^2(i) \quad \text{and} \quad E\left\{N_2^2(i)\right\} = \sigma^2\nu^2(i)$$

We will describe our method first in the case of a 3-D translation parallel to the image plane (panning) and then in the general case of full 3-D translation.

### 3.2. Translation parallel to the image plane

We consider the case where the translational motion along the optical axis is null. According to (1), we can write:

$$\mu(i) = -\frac{T_x f}{Z(i)} \quad \text{and} \quad \nu(i) = -\frac{T_y f}{Z(i)} \qquad (6)$$

Given that the depth $Z(i)$ is unknown, we can only solve for either $\gamma = T_y/T_x$ or $\gamma = T_x/T_y$. This parameter is related to the direction of the translation in the image plane, whose angle to the horizontal axis is given by $\arctan(T_y/T_x)$. We achieve the estimation of this parameter by projecting the observed process on a deterministic process $e(i)$ that is to be specified later. This projection will yield:

$$u_1 = \sum_i u(i)e(i) = \sum_i \mu(i)e(i) + \sum_i N_1(i)e(i) \qquad (7)$$

$$v_1 = \sum_i v(i)e(i) = \sum_i \nu(i)e(i) + \sum_i N_2(i)e(i) \qquad (8)$$

As a consequence of the above assumptions, the mean values of variables $u_1$ and $v_1$ are:

$$E\{u_1\} = -T_x f \sum_i \frac{e(i)}{Z(i)} \quad \text{and} \quad E\{v_1\} = -T_y f \sum_i \frac{e(i)}{Z(i)} \quad (9)$$

Their variances are given by:

$$\mathrm{var}\left\{u_1\right\} = \sigma^2 \sum_i \mu^2(i) e^2(i) \text{and} \mathrm{var}\left\{v_1\right\} = \sigma^2 \sum_i \nu^2(i) e^2(i)$$

We propose to estimate $\gamma = T_y/T_x$ if $u_1 > v_1$, or $\gamma = T_x/T_y$ otherwise. Without loss of generality, we consider the first case, and the estimate will be $\hat{\gamma} = v_1/u_1$.

We will now consider the choice of the axis of projection $\{e(i)\}$. A possible criterion is the maximization of the signal to noise ratio of the denominator variable. This ratio is maximized if $e(i) = \lambda Z(i)$. As $Z(i)$ is unknown but always positive, we propose to choose $e(i) = 1/K$, where $K$ is the number of points. The estimate is then given by:

$$\hat{\gamma} = \frac{\sum_i v(i)}{\sum_i u(i)} \tag{10}$$

Let us now consider the ideal choice $e(i) = \frac{Z(i)}{K}$. We obtain:

$$E\left\{u_1\right\} = -T_x f, \ E\left\{v_1\right\} = -T_y f \tag{11}$$

$$\mathrm{var}\left\{u_1\right\} = \frac{\sigma^2}{K} {T_x}^2 f^2, \ \ \mathrm{var}\left\{v_1\right\} = \frac{\sigma^2}{K} {T_y}^2 f^2 \tag{12}$$

The last equations show the very important reduction of the noise disturbance in estimating $\gamma$, in this ideal case. Indeed, it is known that under the above conditions the estimator is unbiased and efficient, with a variance equal to $\frac{\sigma^2}{K}$. In our case, by selecting $e(i) = 1/K$, the estimator is still unbiased but with a variance proportional to $\frac{\sigma^2}{K}$ with a factor of $\left(1 + \frac{\mathrm{var}(\frac{1}{Z})^2}{(\frac{1}{Z_0})^2}\right)$, where $Z_0$ is the mean depth of the scene. Thus, the efficiency of the estimator depends on the variation of the depth of the scene with respect to its mean value. If the noise is spatially correlated, another factor increases the estimate variance. The stronger the correlation coefficient is, the greater the value of this factor will be. A very important property is that our estimator is unbiased and the associated error is proportional to $\frac{\sigma^2}{K}$.

### 3.3. Translation non parallel to the image plane

We consider the general case where the 3-D translation is not parallel to the image plane ($T_z \neq 0$). We aim at estimating the FOE which is the point $(\alpha, \beta) = (T_x f/T_z, T_y f/T_z)$ in the image plane. We can write:

$$\nu(i)\,\alpha - \mu(i)\,\beta = \nu(i)\,x(i) - \mu(i)\,y(i)$$

From the overdetermined set of equations with noisy coefficients computed from the motion field, we propose to obtain two equations by projecting into two deterministic fields $e_1(i)$ and $e_2(i)$. These two equations are:

$$v_1\,\alpha - u_1\,\beta = w_1 \ \ \text{and} \ \ v_2\,\alpha - u_2\,\beta = w_2 \tag{13}$$

where, for $k = 1, 2$:

$$u_k = \mathbf{u}^T \mathbf{e}_k, \ v_k = \mathbf{v}^T \mathbf{e}_k, \ \text{and} \ w_k = (\mathbf{v}\mathbf{x} - \mathbf{u}\mathbf{y})^T \mathbf{e}_k. \tag{14}$$

We therefore obtain the estimate of the position of the FOE:

$$(\hat{\alpha}, \hat{\beta}) = \left(\frac{u_1 w_2 - u_2 w_1}{u_1 v_2 - u_2 v_1}, \frac{v_2 w_1 - v_1 w_2}{u_1 v_2 - u_2 v_1}\right) \tag{15}$$

We suppose $\sum x(i) = \sum y(i) = \sum x(i) y(i) = 0$. If this is not the case, $x(i)$ and $y(i)$ are expressed in a new coordinate system centered at their centroid and whose orthonormal axes are the first and second principal axes of the distribution of the points. Therefore, if we set $e_1(i) = \lambda\, x(i) Z(i)$ and $e_2(i) = \lambda\, y(i) Z(i)$, we have:

$$E\left\{u_1 v_2 - u_2 v_1\right\} = (\lambda T_z f)^2 \sum_i x^2(i) \sum_i y^2(i) \tag{16}$$

$$E\left\{u_1 w_2 - u_2 w_1\right\} = (\lambda T_z f)^2 \alpha \sum_i x^2(i) \sum_i y^2(i) \tag{17}$$

$$E\left\{v_2 w_1 - v_1 w_2\right\} = (\lambda T_z f)^2 \beta \sum_i x^2(i) \sum_i y^2(i) \tag{18}$$

These relations prove that the proposed estimators are unbiased. Indeed, the quotient (17) / (16) is $\alpha$ and the quantity (18) / (16) is $\beta$. We may also prove that (16) and (17) (idem for (16) and (18)) are decorrelated and that the signal-to-noise ratio for both numerator and denominator is approximately $K$, the number of points. As the depth $Z(i)$ is unknown, we propose to choose as basis $e_1(i) = x(i)$ and $e_2(i) = y(i)$. As in the panning case, the effectiveness of this choice depends on the variation with respect to its mean value, and also on the spatial noise correlation.

## 4. EXPERIMENTAL RESULTS

### 4.1. Results from simulated realistic data

In order to compare the different methods and to study the effect of noise on their accuracy, we use synthetic optical flow fields which are contaminated by different amounts of noise. The simulated optical flow fields are generated using range images from the MSU/WSU Range Image Database, available online at *http://www.eecs.wsu.edu/IRL/RID/*. To simulate a realistic flow field, noise is introduced into the synthetic optical flow vectors. Being similar to the noise model used in [2], we choose the following model, which is compatible with our assumptions:

$$u = \mu + N(0, b) * 0.01 * \mu \ \text{and} \ v = \nu + N(0, b) * 0.01 * \nu \tag{19}$$

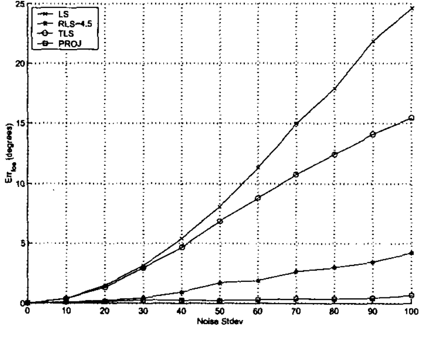

**Fig. 1.** Comparison of the four methods for noise tolerance in the case of full translation

where $N(0,b)$ is a Gaussian random variable with mean 0 and standard deviation $b$. In the case of full translation the assessement criterion is the angular error between the vectors $(\alpha,\beta,f)$ and $(\hat{\alpha},\hat{\beta},f)$, where $(\alpha,\beta)$ is the true FOE and $(\hat{\alpha},\hat{\beta})$ is the estimated one. In the case of panning, the criterion is the error in degrees in the direction of the translation in the image plane.

The methods which have been implemented for comparison are Least Squares (LS), Total Least Squares (TLS), M-estimators Reweighted Least Squares (RLS) and the proposed method (PROJ). These different methods are compared for noise tolerance by computing the angular errors versus the noise standard deviation $b$, varying from 0 to 100%. For each noise level, the computed values are average values over 50 runs. Fig.1 plots the values of angular FOE error in the full translation case. It can be seen from this graph that our method is far more efficient than the other three, giving a maximum error of 0.67 degrees. As expected, LS is the less tolerant to noise. RLS is much more efficient than TLS in the case of full translation. On the other hand, the RLS method being iterative is more time-consuming. In the panning case the conclusions from the experimental results are similar. Our method is even slightly more efficient in this case giving a maximum error of 0.24 degrees.

The proposed method proves to be very tolerant to the noise model we applied which has been found to be close to the one affecting real optical flows. Moreover, it offers the advantage of being very fast and easily implemented, since it consists primarily of projection and summation. In this particular aspect, TLS is more computationaly expensive, performing singular value decomposition.

### 4.2. Results from real data

The algorithms were applied to the well-known "marbled block" and "flower-garden" sequences, with known ground truth values. The "marbled block" sequence was captured by a robot arm moving in full translation over a textured floor. The sequence "flower-garden" corresponds to a panning along the horizontal axis $T_x$ of the camera. The scene contains a tree in the foreground, a textured garden, and a house in the background. "Marbled block" contains many sharp discontinuities in depth and "garden-flower" presents some non-textured areas that cause problem for the optical flow computation, giving rise to a consequent number of outliers. Our method (PROJ) has not been designed explicitly to be optimal in that case. Table 1 gives the results of the different algorithms on these two sequences. The proposed method is the most efficient of the set on these real examples as well, especially in the panning case. These results tend to show that the assumptions made on the noise model and on the criteria of selection of the projection bases were generally valid. As another source for comparison, Daniilidis reported a result with a 7.24° error in FOE for the "Marbled Block" sequence [3].

| **Sequence** | Marbled Block | Flower Garden |
|---|---|---|
| **Type** | Full translation | Panning |
| **Truth** | $(\alpha,\beta)=(-777.0,95.6)$ | $\gamma=0^o$ |
| **LS** | $Err_{\text{FOE}}=7.58^o$ | $Err_\gamma=2.73^o$ |
| **TLS** | $Err_{\text{FOE}}=5.25^o$ | $Err_\gamma=2.67^o$ |
| **RLS** | $Err_{\text{FOE}}=5.42^o$ | $Err_\gamma=2.63^o$ |
| **PROJ** | $Err_{\text{FOE}}=4.94^o$ | $Err_\gamma=1.42^o$ |

**Table 1.** Comparative results on real optical flows

## 5. CONCLUSION

In this paper, we have presented a novel method for estimating the parameters of translational motion from optical flow. Our results on synthetic and real optical flows are more accurate than the other tested methods. This is due to the fact that our scheme, unlike the other methods, is based on an unbiased estimator that strongly reduces the influence of noise contamination in the data. Moreover, computational requirements are low, making this method very attractive for fast 3-D translational motion parameter estimation. We are currently working on the extension of this method to the general case of 3-D motion.

## 6. REFERENCES

[1] D.J Heeger and A.D. Jepson, "Subspace methods for recovering rigid motion i: Algorithm and implementation," *Intern. J. of Computer Vision*, vol. 7, pp. 95–117, 1992.

[2] N. Da Vitoria Lobo and J. K. Tsotsos, "Computing egomotion and detecting independent motion from image motion using collinear points," *Computer Vision and Image Understanding*, vol. 64, pp. 21–52, 1996.

[3] K. Daniilidis, "Fixation simplifies 3d motion estimation," *Computer Vision and Image Understanding*, vol. 68, pp. 158–169, 1997.

[4] N. Komodakis and G. Tziritas, "Robust 3-d motion estimation and depth layering," in *Proc. of Intern. Conf. on Digital Signal Processing*, 1997, pp. 425–428.

[5] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Performance of optical flow techniques," *Intern. J. of Computer Vision*, vol. 12, pp. 43–77, 1994.

[6] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, J. Wiley, 1987.

[7] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image and Vision Computing*, vol. 15, pp. 59–76, 1997.

[8] P. Huber, *Robust statistics*, Wiley, 1981.

[9] E. Simoncelli, E. Adelson, and D. Heeger, "Probability distributions of optical flow," in *Proc. of Computer Vision and Pattern Recognition*, 1991, pp. 310–315.

[10] N. Gupta and L. Kanal, "3-d motion estimation from motion field," *Artificial Intelligence*, vol. 78, pp. 45–86, 1995.

[11] A.M. Earnshaw and S. Blostein, "The performance of camera translation direction estimators from optical flow: Analysis, comparison, and theoretical limits," *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. 18, pp. 927–932, Sept. 1996.