# OPTICAL FLOW ESTIMATION FROM NOISY DATA USING DIFFERENTIAL TECHNIQUES

*Chun-Jen Tsai* *, *Nikolas P. Galatsanos* +, *and Aggelos K. Katsaggelos* *

*Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
tsai,aggk@ece.nwu.edu

+Department of Electrical Engineering
Illinois Institute of Technology, Chicago, IL 60616, USA
npg@ece.iit.edu

**ABSTRACT**

Many optical flow estimation techniques are based on the differential optical flow equation. These algorithms involve solving over-determined systems of optical flow equations. Least squares (LS) estimation is usually used to solve these systems even though the underlying noise does not conform to the model implied by LS estimation. To ameliorate this problem, work has been done using the total least squares (TLS) method instead. However, the noise model presumed by TLS is again different from the noise present in the system of optical flow equations. A proper way to solve the system of optical flow equation is the constrained total least squares (CTLS) technique. The derivation and analysis of the CTLS technique for optical flow estimation is presented in this paper. It is shown that CTLS outperforms TLS and LS optical flow estimation.

## 1. INTRODUCTION

There are many methods for optical flow estimation for a single channel image sequence. For techniques based on the differential optical flow equation, a locally constant flow model is usually used to allow the construction of an over-determined system of constraint equations, $A\mathbf{x} = \mathbf{b}$, where $A$ is composed of spatial intensity derivatives and **b** contains the temporal intensity derivatives. An estimate of the optical flow vector **x** is then obtained using least squares (LS) estimation [1]. LS fit produces robust, but not very accurate estimates, since the LS model does not account for the noise in the spatial derivative matrix $A$.

Total least squares (TLS) is a technique which takes into account the noise in $A$ [2, 3]. Some work has been done on optical flow estimation using TLS techniques [4]-[9]. In [4], TLS is used to replace LS directly for optical flow fitting, assuming a locally smooth flow field model. Without extra effort to regularize the estimated field, TLS tends to give noisier estimates [7, 10] when the i.i.d. noise model does not hold or when highly inconsistent systems of equations are encountered. In [4], rank-deficient cases are detected by comparing intensity gradient directions in a neighborhood and treating each of them differently, while in [9] a reliability measure from the singular value decomposition of $[A|\mathbf{b}]$ is used to regularize the estimated field.

Although the regularized TLS approach gives more accurate estimates than the LS technique, improvements can be expected if a correlated noise model is used [10, 11, 12]. An extension to TLS which takes into account the noise correlation in $[A|\mathbf{b}]$ is the constrained total least squares (CTLS) method. CTLS has been successfully used to solve image restoration problems [10, 11]. In [7], an error-in-variable (EIV) formulation of optical flow estimation is presented. This approach can be shown to be equivalent to the CTLS approach [13]. In this paper, the CTLS approach is used to solve the optical flow estimation problem.

The paper is organized as follows. In section 2, optical flow estimation using a first order differential constraint is introduced. Section 3 derives the nonlinear cost function for the CTLS approach. Section 4 presents a performance analysis that compares different techniques with the Cramer-Rao lower bound. Experiments and conclusions are presented in sections 5 and 6, respectively.

## 2. PROBLEM FORMULATION

The most commonly used constraint in optical flow estimation is the optical flow equation:

$$\frac{\partial E_t}{\partial x}\frac{dx}{dt} + \frac{\partial E_t}{\partial y}\frac{dy}{dt} + \frac{\partial E_t}{\partial t} = 0, \tag{1}$$

where $E_t(\cdot)$ is the image intensity at time $t$, $\partial E_t/\partial x$ and $\partial E_t/\partial y$ are the spatial derivatives of the image intensity function $E_t(\cdot)$, $\partial E_t/\partial t$ is the temporal derivative of the image intensity function, and $(dx/dt, dy/dt)^T$ is the optical flow vector. Least squares (LS) estimation is often used to compute the optical flow. With a square $\sqrt{m} \times \sqrt{m}$ estimation window, the resulting problem is the solution of the following over-determined system of equations

$$A_{m\times 2}\ \mathbf{x} = \mathbf{b}_{m\times 1} \tag{2}$$

where

$$A = \begin{pmatrix} \frac{\partial E_t}{\partial x}(\mathbf{s}_1) & \frac{\partial E_t}{\partial y}(\mathbf{s}_1) \\ & \ldots \\ \frac{\partial E_t}{\partial x}(\mathbf{s}_m) & \frac{\partial E_t}{\partial y}(\mathbf{s}_m) \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} -\frac{\partial E_t}{\partial t}(\mathbf{s}_1) \\ \ldots \\ -\frac{\partial E_t}{\partial t}(\mathbf{s}_m) \end{pmatrix},$$

with $\mathbf{x} = (u, v)^T$ being the optical flow vector at image positions $\mathbf{s}_i = (x_i, y_i)^T, i = 1, ..., m$. If neither $A$ nor $\mathbf{b}$ contain noise, Eq. (2) should be a consistent system and have an exact solution. However, in optical flow estimation, $A$ is composed of the spatial gradients, which are estimated using numerical differentiation of the noisy image data. Better performance can be expected if the noise in $A$ and the fact that correlation exists between gradient estimates of neighboring pixels is taken into account [12].

## 3. CONSTRAINED TOTAL LEAST SQUARES FOR OPTICAL FLOW ESTIMATION

Let $A$ and $\mathbf{b}$ the noisy matrices for the system of optical flow equations as defined in Eq. (2), and $\Delta A$ and $\Delta\mathbf{b}$ be the perturbation matrices that cancel out the noise in $A$ and $\mathbf{b}$, respectively. That is,

$$(A + \Delta A)\mathbf{x} = \mathbf{b} + \Delta\mathbf{b},$$

or equivalently,

$$A\mathbf{x} + (\Delta A\mathbf{x} - \Delta\mathbf{b}) = \mathbf{b}. \tag{3}$$

If the noise in $\Delta A$ and $\Delta\mathbf{b}$ can be modeled as the multiplication of a coloring matrix $L(\mathbf{x})$ with a white noise vector $\epsilon$, we can reformulate the estimation problem using CTLS. That is,

$$\Delta A\mathbf{x} - \Delta\mathbf{b} = L(\mathbf{x}) \cdot \epsilon. \tag{4}$$

The problem then takes the form:

$$\min \|\epsilon\|_2, \quad \text{subject to } \ A\mathbf{x} + L(\mathbf{x}) \cdot \epsilon = \mathbf{b}. \tag{5}$$

Solving for $\epsilon$ from the constraint in Eq. (5) leads to $\epsilon = L^{+}(\mathbf{x})\,(\mathbf{b} - A\mathbf{x})$, where $L^{+}(\mathbf{x})$ denotes the Moore-Penrose pseudo inverse of $L(\mathbf{x})$. The minimization of $\|\epsilon\|_2$ is equivalent to the minimization of $\epsilon^T\epsilon$. The cost function can then be defined as follows:

$$J(\mathbf{x}) = (\mathbf{b} - A\mathbf{x})^T L^{+}(\mathbf{x})^T L^{+}(\mathbf{x})\,(\mathbf{b} - A\mathbf{x})\,. \tag{6}$$

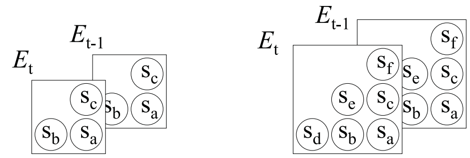


Figure 1: Left: the neighborhood structure for derivative estimation. Right: pixels used for optical flow estimation at position $\mathbf{s}_a$.

To apply this technique to optical flow estimation, we must define the noise vector $\epsilon$ and derive the coloring matrix $L(\mathbf{x})$ according to the noise model in the matrices $A$ and $\mathbf{b}$.

The spatio-temporal derivatives $\frac{\partial E_t}{\partial x}(\mathbf{s}_i)$, $\frac{\partial E_t}{\partial y}(\mathbf{s}_i)$, and $\frac{\partial E_t}{\partial t}(\mathbf{s}_i)$ are usually estimated using finite difference equations. Without loss of generality and for ease of presentation, $L(\mathbf{x})$ is derived in the following paragraph using a simple two-point backward difference equation. Given the neighborhood structure shown in Figure 1, the image intensity derivatives at pixel position $\mathbf{s}_a$ can be calculated using the following equations:

$$\begin{cases} \frac{\partial E_t}{\partial x}(\mathbf{s}_a) = E_t(\mathbf{s}_a) - E_t(\mathbf{s}_b) \\ \frac{\partial E_t}{\partial y}(\mathbf{s}_a) = E_t(\mathbf{s}_a) - E_t(\mathbf{s}_c) \\ \frac{\partial E_t}{\partial t}(\mathbf{s}_a) = E_t(\mathbf{s}_a) - E_{t-1}(\mathbf{s}_a) \end{cases} \tag{7}$$

If the image intensities are corrupted by i.i.d. noise, the entries in $A$ and $\mathbf{b}$ would be corrupted by correlated noise. If a 3-point estimation window (Figure 1) and Eq. (7) are used, $A$ and $\mathbf{b}$ can be written as:

$$\begin{aligned} A &= \begin{pmatrix} E_t(\mathbf{s}_c) - E_t(\mathbf{s}_e) & E_t(\mathbf{s}_c) - E_t(\mathbf{s}_f) \\ E_t(\mathbf{s}_b) - E_t(\mathbf{s}_d) & E_t(\mathbf{s}_b) - E_t(\mathbf{s}_e) \\ E_t(\mathbf{s}_a) - E_t(\mathbf{s}_b) & E_t(\mathbf{s}_a) - E_t(\mathbf{s}_c) \end{pmatrix} \text{ and} \\ \mathbf{b} &= -\begin{pmatrix} E_t(\mathbf{s}_c) - E_{t-1}(\mathbf{s}_c) \\ E_t(\mathbf{s}_b) - E_{t-1}(\mathbf{s}_b) \\ E_t(\mathbf{s}_a) - E_{t-1}(\mathbf{s}_a) \end{pmatrix}. \end{aligned} \tag{8}$$

We assume that the image intensities are corrupted by i.i.d. noise, i.e.

$$\begin{aligned} E_t(\mathbf{s}_j) &= \bar{E}_t(\mathbf{s}_j) + \epsilon_j, \ \ j = a, b, c, d, e, f \ \ \textit{and} \\ E_{t-1}(\mathbf{s}_k) &= \bar{E}_{t-1}(\mathbf{s}_k) + \epsilon'_k, \ \ k = a, b, c, \end{aligned} \tag{9}$$

where $\bar{E}_t(\cdot)$ is the true image intensity at time $t$, and $\epsilon_a, ..., \epsilon_f$ and $\epsilon'_a, \epsilon'_b, \epsilon'_c$ are i.i.d. zero mean noise with variance $\sigma_\epsilon^2$. Let $\mathbf{x} = (u\ v)^T$, and

$$\epsilon = \begin{pmatrix} \epsilon_d & \epsilon_e & \epsilon_f & \epsilon_c & \epsilon_b & \epsilon_a & \epsilon'_c & \epsilon'_b & \epsilon'_a \end{pmatrix}^T.$$

We can define $L(\mathbf{x})$ as

$$L(\mathbf{x}) = \begin{pmatrix} 0 & -v & -u & u+v+1 \\ -v & -u & 0 & 0 \\ 0 & 0 & 0 & -u \end{pmatrix} \tag{10}$$

$$\begin{array}{ccccc} 0 & 0 & -1 & 0 & 0 \\ u+v+1 & 0 & 0 & -1 & 0 \\ -v & u+v+1 & 0 & 0 & -1 \end{array} \Bigg)_{3\times 9} \sigma_\epsilon$$

Since $L(\mathbf{x})$ is a full rank matrix with more columns than rows, $L^+(\mathbf{x}) = L(\mathbf{x})^T \left(L(\mathbf{x})L(\mathbf{x})^T\right)^{-1}$. The noise covariance matrix $\Sigma(\mathbf{x})$ can be computed as

$$\Sigma(\mathbf{x}) = \left(L^+(\mathbf{x})^T L^+(\mathbf{x})\right)^{-1} = \begin{pmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_3^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_4^2 \\ \sigma_3^3 & \sigma_4^2 & \sigma_1^2 \end{pmatrix} \sigma_\epsilon^2, \tag{11}$$

where

$$\begin{aligned} \sigma_1^2 &= v^2 + u^2 + (u+v+1)^2 + 1 \\ \sigma_2^2 &= uv \\ \sigma_3^2 &= -v(u+v+1) \\ \sigma_4^2 &= -u(u+v+1). \end{aligned}$$

The minimization of $J(\mathbf{x})$ in Eq. (6) with respect to $\mathbf{x}$ results in the CTLS estimate of the optical flow at pixel position $\mathbf{s}_a$. This method can be generalized for larger window sizes and for different finite difference equations.

## 4. PERFORMANCE BOUND

It is very insightful to compare the variance of an unbiased estimator to the theoretical lower bound of the variance of the estimator based on the same noise model. The most widely quoted variance bound is probably the Cramer-Rao lower bound (CRLB). Assuming a general Gaussian noise model for the observation vector, i.e.,

$$\begin{aligned} A\mathbf{x} &= \mathbf{b}, \;\; \mathbf{x} = (u, v, )^T \\ p(\mathbf{b}|\mathbf{x}) &= N(A\mathbf{x}, \Sigma(\mathbf{x})), \end{aligned}$$

the CRLB is determined by the following equation

$$var(x_i) \geq \left[I^{-1}(\mathbf{x})\right]_{ii}, \tag{12}$$

where $x_i$ is the $i$th component of the unknown vector $\mathbf{x}$ and, $I(\mathbf{x})$ is the Fisher information matrix, whose $(i,j)$th elements is given by

$$\begin{aligned} [I(\mathbf{x})]_{ij} &= -E\left[\frac{\partial^2 \log p(\mathbf{b}|\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j}\right] \\ &= \left[\frac{\partial A\mathbf{x}}{\partial x_i}\right]^T \Sigma(\mathbf{x})^{-1} \left[\frac{\partial A\mathbf{x}}{\partial x_j}\right] \\ &+ \frac{1}{2} tr\left[\Sigma(\mathbf{x})^{-1}\frac{\partial \Sigma(\mathbf{x})}{\partial x_i}\Sigma(\mathbf{x})^{-1}\frac{\partial \Sigma(\mathbf{x})}{\partial x_j}\right] \end{aligned} \tag{13}$$

with $tr\,[A]$ the trace of matrix A. For the details of the computation of $\frac{\partial A\mathbf{x}}{\partial x_i}$, please refer to [13].

Figure 2: A frame from a simulated image sequence.

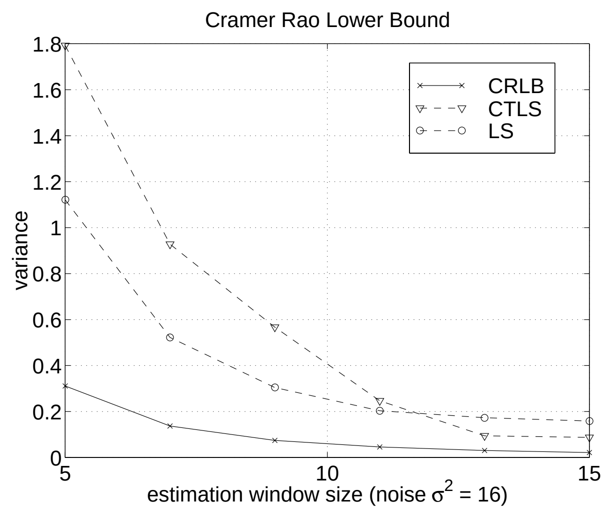


Figure 3: Variance of estimates as a function of window size

## 5. EXPERIMENT

In this section simulations are conducted first to show the accuracy of LS and CTLS methods as a function of the estimation window size. In the simulation, a sequence of a single object undergoing uniform motion is synthesized. The simulated sequence is generated by translating the $256 \times 256$-pixel image shown in Figure 2 to simulate a 2-pixel motion in the horizontal direction. There is no motion in the vertical direction. The image sequence is corrupted by a zero mean white Gaussian noise with variance $\sigma_\epsilon^2 = 16$. The optical flow estimation is performed on 64 pixel sites in the center of the images. The variances of the estimates are then averaged over these pixel sites.

Figure 3 shows the estimator variance versus the estimation window size. The y-axis is the sum of variances of the optical flow estimates $\hat{u}$ and $\hat{v}$. From this figure, one can see that CTLS is less accurate when the estimation window size is small. This is probably because the CTLS technique needs a large number of observations to capture the noise statistics. As the number of observations gets larger, the estimates from CTLS become more accurate and eventually, CTLS out-performs LS.

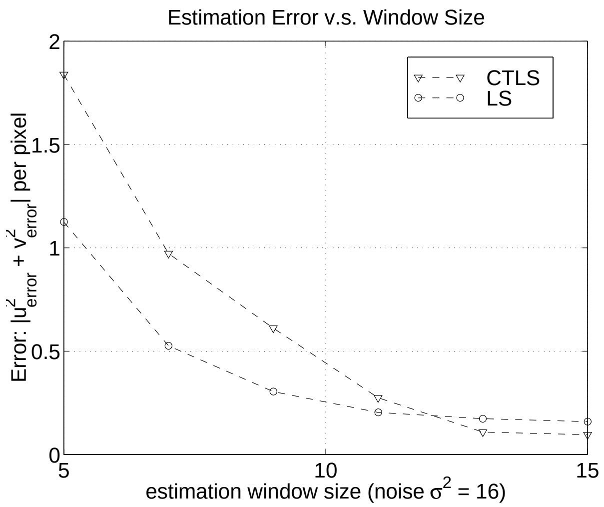


Figure 4: Estimation error as a function of window size

| *window size* | *CTLS* |
|---|---|
| $7 \times 7$ | 35 sec. |
| $11 \times 11$ | 486 sec. |
| $15 \times 15$ | 1130 sec. |

Table 1: Computation time versus window size.

Using the same simulated image sequence, the mean squared error (MSE) of the estimates versus estimation window size is shown in Figure 4. Figure 4 is very similar to Figure 3 because CTLS and LS are unbiased estimators. In this case, the MSE should be the same as the variance of the estimates. The computation time on a Pentium II 400MHz machine running Linux operating system is listed in Table 1. Note that the computation time depends on the nonlinear optimization algorithm used to solve Eq. (6). Both conjugate gradient method and Powell's conjugate direction method have been tested. The results shown are from the conjugate gradient method.

## 6. CONCLUSIONS

According to the experiments, the CTLS optical flow estimation technique applied to single channel video sequences outperforms the LS technique when the estimation window size is large. The major disadvantage of the CTLS technique is its high computational cost. For example, when a $15 \times 15$ window is used, each evaluation of the optimization process involves the computation of the pseudo inverse of a $255 \times 255$ matrix. To make this technique practical, one would prefer to keep the estimation window size under $11 \times 11$. We are investigating the application of other constraints to improve the performance of CTLS using small estimation windows.

## 7. REFERENCES


[1] B. Lucas and T. Kanade, "An Iterative image registration technique with an application to stereo vision," *DARPA Image Understanding Workshop*, pp. 121-130, 1981.

[2] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numerical Analysis*, **17** pp. 883-893, 1980.

[3] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem*, SIAM, Philadelphia, 1991.

[4] C. H. Chu and E. J. Delp, "Estimating Displacement Vectors from an Image Sequence," *Journal of Opt. Soc. Am. A*, vol. 6, no. 6, pp. 871-878, June 1989.

[5] S. Wang, V. Markandey, and A. Reid, "Total least squares fitting spatiotemporal derivatives to smooth optical flow fields," *Proc. SPIE: Signal and Data Processing of Small Targets*, vol. 1698, pp. 42-55, 1992.

[6] J. Weber and J. Malik, "Robust Computation of Optical Flow in a Multi-scale differential framework," *Int. J. of Computer Vision*, vol. 14, pp. 67-81, 1995.

[7] L. Ng and V. Solo, "Error-in-variable Modeling in Optical Flow Problems," *Proc. ICASSP*, Seattle, Washinton, pp. 2773-2776, 1998.

[8] Vania Estrela, *Motion Estimation Using the Total Least Squares Approach.* Dept. of ECE, M.S. thesis, Northwestern University, Dec. 1995

[9] C.-J. Tsai, N. Galatsanos and A. K. Katsaggelos, "Total Least-Squares Estimation of Stereo Optical Flow," to appear in *Proc. IEEE Int. Conf. on Image Processing*, Chicago, Oct., 1998.

[10] T.J. Abatzoglou, J.M. Mendel, G.A. Harada, "The Constrained Total Least Squares Technique and Its Application to Harmonic Superresolution," *IEEE Trans. Signal Processing* vol. 39, No. 5, pp. 1070-1086, May 1991.

[11] V. Z. Mesarovic, N. P. Galatsanos and A. K. Katsaggelos, "Regularized constrained total least-squares image restoration," *IEEE Trans. Image Processing*, vol. 4, no. 8, pp. 1096-1108, Aug. 1995.

[12] H. Nagel, "Optical Flow Estimation and the Interaction between Measurement Errors at Adjacent Pixel Positions," *Int. J. of Computer Vision*, vol. 15, pp. 271-288, 1995.

[13] C.-J. Tsai, *Disparity and Motion Fields Estimation in Bi-channel Video Sequences*, Ph.D. Dissertation, Northwestern University, Dept. of ECE, Dec., 1998.