MAXIMUM-LIKELIHOOD OPTICAL FLOW ESTIMATION USING DIFFERENTIAL CONSTRAINTS

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. For these techniques, a locally constant flow model is typically used to allow the construction of an over-determined system of constraint equations. In this paper, the problem of solving the system of optical flow equations using a constraint total least squares (CTLS) approach is investigated. It is shown that by modifying the CTLS approach it becomes identical to a maximum likelihood (ML) approach to the problem. This modification improves the CTLS estimates especially when the estimation window size is small, as is demonstrated experimentally.

1. INTRODUCTION

Optical flow estimation plays an important role in many visual communication applications. There are a number of techniques reported in the literature for the solution of this problem [1]. When a video sequence with high frame rate is considered, differential techniques are of particular interest. For techniques based on the *optical flow equation* [1], a locally constant flow model is typically used to allow for the construction of an over-determined system of constraint equations, $A\mathbf{x} = \mathbf{b}$, where A is composed of spatial intensity derivatives, $\mathbf{b}$ contains the temporal intensity derivatives, and $\mathbf{x}$ represents the optical flow field. In this formulation, A and $\mathbf{b}$ are corrupted by correlated noise.

Some work has been done in dealing with correlated noise in image processing [4, 2, 5, 3]. For example, the total least squares (TLS) method has been extended to the constrained total least squares (CTLS) method by taking into account the noise correlation in $[A|\mathbf{b}]$. It has been successfully used to solve image restoration problems [4, 2]. However, when applying this technique to optical flow estimation, it tends to give noisy results when the estimation window size is small [7]. Although thresholding can be used to regularize the estimation process using a reliability measure, an alternative way is proposed in this paper to solve the problem of noisy estimates of CTLS when the number of observations is small.

There are two major sources of errors in the system of optical flow equations. The first one is due to the noise that corrupts the image and the second one is the error introduced by the numerical approximation of the derivative estimates. Even though CTLS addresses one of the problems in the system of optical flow equations, namely, the fact that both A and $\mathbf{b}$ are corrupted by correlated noise, it does not take into account the fact that in general the noise which corrupts $\mathbf{b}$ (composed of temporal derivatives) is larger than the noise which corrupts A (composed of spatial derivatives). Since total least squares based techniques are very sensitive to noise model, CTLS tends to give noisy results when the number of observations is small.

In this paper, we introduce an adjustment term to the cost function derived from CTLS. The resulting cost function is equivalent to a maximum likelihood (ML) formulation for optical flow estimation. Experiments show that this new technique performs better than CTLS, especially when the window size is small.

2. PROBLEM FORMULATION

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

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

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. Using a square $\sqrt{m} \times \sqrt{m}$ estimation window, the optical flow estimation problem can be expressed as the solution of an over-determined system of optical flow equations [8]

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

where

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

with $\mathbf{x} = (u, v)^T$ being the optical flow vector at image positions $\mathbf{s}_i = (x_i, y_i)^T, i = 1, \dots, m$. If neither A nor $\mathbf{b}$ are corrupted by 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 [5].

3. CTLS FORMULATION

Let ΔA and $\Delta\mathbf{b}$ be the perturbation matrices that result in a consistent system of equations, 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 ΔA and $\Delta\mathbf{b}$ can be modeled by

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

where $L(\mathbf{x})$ is a coloring matrix and ϵ a white noise vector, the problem takes the form:

$$\min \|\epsilon\|_2^2, \quad \text{subject to } A\mathbf{x} + L(\mathbf{x}) \cdot \epsilon = \mathbf{b}. \tag{5}$$

Solving for ϵ 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 $\|\epsilon\|_2^2$ then results to the minimization of

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

where $\Sigma(\mathbf{x})^+ = L^+(\mathbf{x})^T L^+(\mathbf{x})$. To apply this technique to the estimation of the optical flow, we must define the noise vector ϵ and derive the coloring matrix $L(\mathbf{x})$ according to the noise model in the spatio-temporal matrix $[A|\mathbf{b}]$.

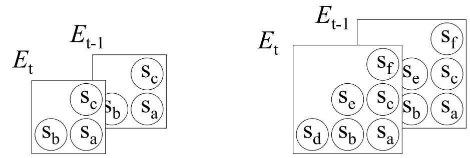

Figure 1: Left: the neighborhood structure for derivative estimation. Right: pixels used for optical flow estimation at position $\mathbf{s}_a$.

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 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 a 3-point estimation window (Figure 1) and Eq. (7) are used, A and $\mathbf{b}$ can be written as:

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

If the image intensities are corrupted by i.i.d. noise, the entries in A and $\mathbf{b}$ are corrupted by correlated noise. That is,

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

where $\bar{E}_t(\cdot)$ is the true image intensity at time t, and $\epsilon_a, \dots, \epsilon_f$ and $\epsilon'_a, \epsilon'_b, \epsilon'_c$ represent i.i.d. zero mean noise with variance σ_ϵ^2. Letting

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

the matrix $\Sigma(\mathbf{x})$ can be computed as ([7])

$$\Sigma(\mathbf{x}) = \left(L^+(\mathbf{x})^T L^+(\mathbf{x})\right)^+ = \begin{pmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_3^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_4^2 \\ \sigma_3^3 & \sigma_4^2 & \sigma_1^2 \end{pmatrix} \sigma_\epsilon^2, \tag{10}$$

where

$$\sigma_1^2 = v^2 + u^2 + (u + v + 1)^2 + 1$$

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

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. ML FORMULATION

In previous discussions, a general probability distribution is assumed for the noise vector ϵ. When ϵ is a correlated Gaussian noise vector, we have

$$\mathbf{b} = A\mathbf{x} + \mathbf{n}, \ \ \mathbf{n} \sim N(0, C(\mathbf{x})), \tag{11}$$

where $C(\mathbf{x})$ is the noise covariance matrix. The probability density function of $p(\mathbf{b}|\mathbf{x})$ can be written as

$$\begin{aligned} p(\mathbf{b}|\mathbf{x}) &= N(A\mathbf{x}, C(\mathbf{x})) \\ &= \frac{1}{(2\pi)^{N/2}} |C(\mathbf{x})|^{-1/2} \times \\ &\quad \exp\left\{ \frac{-1}{2} (A\mathbf{x} - \mathbf{b})^T \ C(\mathbf{x})^{-1} (A\mathbf{x} - \mathbf{b}) \right\}. \end{aligned} \tag{12}$$

The ML estimator can be elaborated as follows:

$$\begin{aligned} \mathbf{x}_{ML} &= \arg\max_{\mathbf{x}} \{p(\mathbf{b}|\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}} \{-\log p(\mathbf{b}|\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}} \left\{ \frac{N}{2} \log(2\pi) + \frac{1}{2} \log |C(\mathbf{x})| \right. \\ &\quad \left. + \frac{1}{2} (A\mathbf{x} - \mathbf{b})^T C(\mathbf{x})^{-1} (A\mathbf{x} - \mathbf{b}) \right\} \\ &= \arg\min_{\mathbf{x}} \{\log |C(\mathbf{x})| + \\ &\quad \left. (A\mathbf{x} - \mathbf{b})^T \ C(\mathbf{x})^{-1} (A\mathbf{x} - \mathbf{b}) \right\}. \end{aligned} \tag{13}$$

Therefore, the cost function can be defined as:

$$J_{ML}(\mathbf{x}) = \log |C(\mathbf{x})| + (\mathbf{b} - A\mathbf{x})^T \ C(\mathbf{x})^{-1} (\mathbf{b} - A\mathbf{x}). \tag{14}$$

$\log |C(\mathbf{x})|$ can be ignored when Eq. (13) is used to solve estimation problems where the observation vectors are large. However, for optical flow estimation, since the observation vector is relatively small (25-by-1 for a 5×5 estimation window), $\log |C(\mathbf{x})|$ has significant contribution to the error function. In Eq. (6), $\Sigma(\mathbf{x})$ plays a similar role as $C(\mathbf{x})$ in Eq. (14). It is shown empirically that by adding $\log |\Sigma(\mathbf{x})|$ to Eq. (6), the performance of the CTLS optical flow estimator is improved.

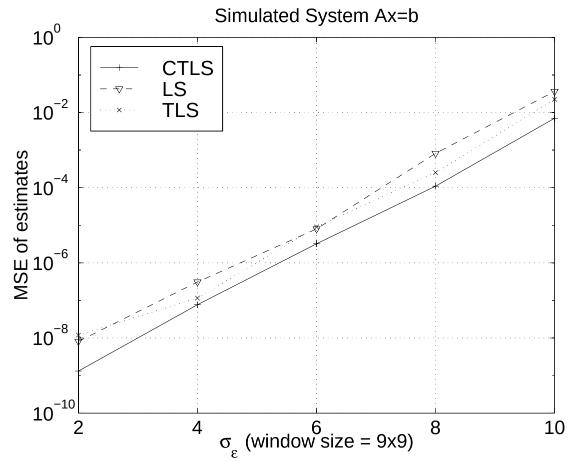

Figure 2: The performance of CTLS when the model fits.

5. EXPERIMENT

Total least squares based techniques are very sensitive to model errors. In this section, a simulation is first conducted to show the performance of the CTLS estimator when the noise model is accurate. In this experiment, Eq. (4) is used to generate a system of optical flow equations $A\mathbf{x} = \mathbf{b}$. Here, ϵ is a 50×1 random vector composed of i.i.d. Gaussian noise components. The size of the estimation window is 9×9. The coloring matrix $L(\mathbf{x})$ is a linear function of $\mathbf{x}$. That is, $L(\mathbf{x}) = [L_1|L_2]\mathbf{x} + L_3$, where L_1, L_2, and L_3 are constant 81×50 matrices. Given ϵ and $L(\mathbf{x})$, ΔA and $\Delta\mathbf{b}$ can be generated to form Eq. (3). The true solution of $\mathbf{x}$ is the vector $[1\ 1]^T$. Three different techniques, LS, TLS, and CTLS are used to estimate $\mathbf{x}$ and the mean squared errors (MSE) of the estimates versus different noise levels are shown in Fig. 2. In this example, the components in A and $\mathbf{b}$ conform to the noise model assumption of the CTLS approach. The estimates from the LS and CTLS methods are not thresholded while the estimates from the TLS method are thresholded using the reliability measure to be discussed later.

In practical situations, unless a very high frame rate image sequence is used, the noise that corrupts $\mathbf{b}$ is usually much larger than the noise that corrupts A due to the difficulty in temporal derivative estimation. Therefore, CTLS does not perform well and thresholding is required to regularize the estimates [7]. By introducing the extra term from the ML formulation, the estimates become more accurate even without thresholding. In the second and third experiments, a sequence of an image undergoing uniform motion is synthesized. The simulated sequence is generated by translating the 256×256-pixel “USC woman” image to simulate a 2-pixel motion in the horizontal direction. There is no motion in the vertical direction.

Fig. 3 shows the performance of the estimator with different window sizes. The image sequence is corrupted by zero mean white Gaussian noise with variance $\sigma_\epsilon^2 = 16$. For the CTLS technique, singular value decomposition of $[A|\mathbf{b}]$ is first computed for thresholding purpose. The right singular vector, α, associated with the smallest singular value can be used as a reliability measure. If $\alpha = [\alpha_1, \alpha_2, \alpha_3]^T$, the estimate is considered unreliable when α_3 is less than a predetermined threshold value [6]. The ML estimates are not thresholded. It is clear that when the window size is small, ML performs better than CTLS estimation.

The third experiments shows the performance of the estimator under different noise levels. In Fig. 4, the image sequence is corrupted by white Gaussian noise and a 5×5 window is used for estimation. Notice that ML estimation does not need thresholding and it performs better than thresholded CTLS estimation. As the experiments show, adding the extra term $\log|\Sigma(\mathbf{x})|$ to $J(\mathbf{x})$ improves the performance of the CTLS estimator.

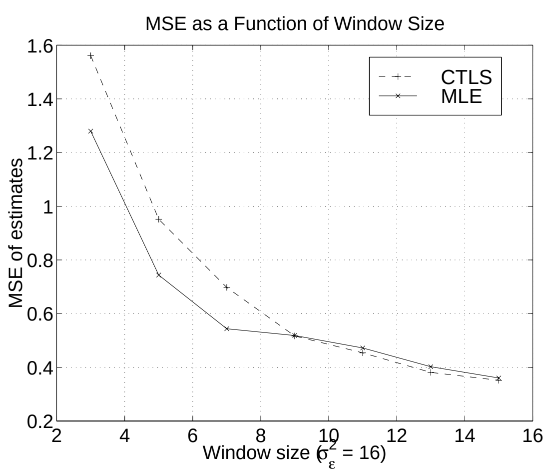

Figure 3: Estimation error as a function of window size.

6. CONCLUSIONS

According to the experiments, the additional term in the ML formulation helps the CTLS approach to get better results when the number of observations is small. This is important because the CTLS method has high computational cost when the estimation window is large. Another problem with the CTLS technique is that it is very sensitive to the noise model. In the case of optical flow estimation, the noise in $\mathbf{b}$ is typically much higher than the noise in A. This deviation from the assumption that A and $\mathbf{b}$ are corrupted by noise of similar strength hinders the performance of the CTLS method even with the additional correction term. Possible solutions to this problem are currently under investigation.

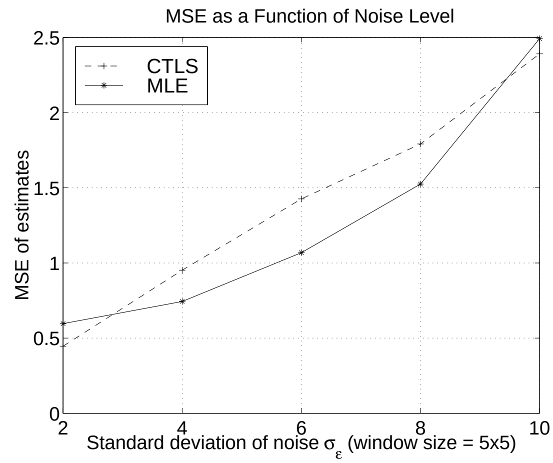

Figure 4: Estimation error as a function of noise level.

7. REFERENCES

[1] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "System and Experiment Performance of Optical Flow Techniques," *Int. J. of Computer Vision*, Vol. 12:1, pp. 43-77, 1994.

[2] 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.

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

[4] 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.

[5] 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.

[6] C.-J. Tsai, N.P. Galatsanos and A. K. Katsaggelos, "Total Least-Squares Estimation of Stereo Optical Flow," *Proc. IEEE Int. Conf. on Image Processing*, Vol. II, pp. 622-626, Chicago, Oct., 1998.

[7] C.-J. Tsai, N.P. Galatsanos and A. K. Katsaggelos, "Optical Flow Estimation from Noisy Data Using Differential Techniques," *Proc. ICASSP'99*, p.VI-3393, Phoenix, AZ, Mar. 1999.

[8] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, no. 1-3, pp. 185-204, 1981.