AN ADAPTIVE SCHEME FOR ESTIMATING MOTION

Hassan Foroosh

School of Computer Science
University of Central Florida
Orlando, FL, 32816-3262
foroosh@cs.ucf.edu

ABSTRACT

In this paper, we propose a new scheme for estimating the motion adaptively at each pixel. We assume non-stationary statistics for noise and motion, and propose a blind approach based on using the Generalized Cross Validation (GCV) technique. We thus determine a different level of regularization at each pixel based on the local statistics using GCV. The problems addressed include the preservation of discontinuities, model/data errors, and confidence measures. Experimental results demonstrate the high accuracy of the prposed technique applied to both synthetic and real data.

1. INTRODUCTION

The literature on motion analysis is quite rich and covers over two decades of research in image processing and computer vision. The focus of this paper is the gradient-based techniques, which were motivated in by the work of Horn and Schunk [6]. The underlying assumption in all existing contributions [1, 2, 5, 6, 7, 8] is the intensity conservation along the trajectory of pixels, which yields the well-known gradient constraint equation

$$f_x u + f_y v + f_t = 0 \qquad (1)$$

where $f(x, y, t)$ is the image sequence, the subscripts indicate the partial derivatives with respect to the corresponding variables, and (u, v) is the vector of two unknown components of the velocity at each pixel.

Clearly, the problem is underconstrained - one linear equation and two unknowns. Two approaches that attempt to tackle this issue are based on the use of neighborhood information (piecewise constancy or smoothness), or the use of higher order derivatives [7, 10]. A major issue in motion estimation is errors in model (e.g. oversmoothing at discontinuities) or the noise on the data, which we have addressed in this paper.

This work was partially supported by ONR grant #N00014-04-1-0512, and Sun Microsystem's grant #EDUD-7824-030482-US.

2. PROPOSED APPROACH

Let us rewrite the gradient constraint equation as follows:

$$\nabla^T f \tilde{\mathbf{u}} = 0 \qquad (2)$$

where $\nabla = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t}\right]^T$ is the usual gradient operator, $\tilde{\mathbf{u}} = [u, v, 1]^T$ is a homogenous representation of the velocity vector, and $\nabla^T f$ is a shorthand for $(\nabla f)^T$, where the superscript T denotes the transposition.

Applying the gradient operator to both sides yields

$$\nabla\nabla^T f \tilde{\mathbf{u}} = \mathcal{H}\tilde{\mathbf{u}} = \mathbf{0} \qquad (3)$$

where $\mathcal{H}$ is the 3×3 Hessian matrix of $f(x, y, t)$ at each pixel. This is a system of linear simultaneous equations, where the first two equations lead essentially to the formulation proposed by Uras et al. [10], and the third equation is a new temporal constraint [3], which implies that either there is no acceleration or the temporal sampling of the sequence is sufficiently high so that inter-frame accelerations is negligible.

If we now explicitly enforce the gradient constraint equation in (3), we obtain the following model to solve for the motion field

$$\begin{bmatrix} \nabla\nabla^T f \\ \nabla^T f \end{bmatrix} \tilde{\mathbf{u}} = \tilde{\mathbf{H}}\tilde{\mathbf{u}} = \mathbf{0} \qquad (4)$$

In practice due to various sources of error equation (4) does not identically vanish. Hence, we get

$$\tilde{\mathbf{H}}\tilde{\mathbf{u}} = \mathbf{e} \qquad (5)$$

where $\mathbf{e}$ is a random vector whose statistics are unknown.

Although one could solve (5) simply by using least-squares, the results would be affected by various sources of inaccuracy such as model or data errors, motion discontinuities and the aperture problem. Therefore in this section, we propose an adaptive approach based on GCV. The adaptiveness of our scheme is due to two factors. One is

0-7803-8554-3/04/$20.00 ©2004 IEEE.

our non-stationarity assumption, which leads to a spatially varying regularization parameter computed independently at each pixel, and the other is the use of a norm regularization described below.

Use of GCV in the context of motion estimation has never been exploited before. We particularly derive a closed-form expression for the optimal regularization parameters that has not been reported in the past, and that lends itself to an easy implementation.

Let us rewrite (5) as follows

$$\mathbf{H}\mathbf{u} = \mathbf{b} + \mathbf{e} \tag{6}$$

where $\mathbf{H}$ is a 4×2 matrix given by $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2]$, and $\mathbf{b} = -\mathbf{h}_3$, with $\mathbf{h}_1, ..., \mathbf{h}_3$ denoting the columns of $\tilde{\mathbf{H}}$.

A classical remedy for overcoming the errors due to $\mathbf{e}$ is given by the following generalization of Tikhonov regularization

$$\hat{\mathbf{u}} = (\mathbf{H}^T\mathbf{H} + \mu_{xy}\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{b} \tag{7}$$

which minimizes

$$Q(\mathbf{u}) = \|\mathbf{H}\mathbf{u} - \mathbf{b}\|^2 + \mu_{xy}\|\mathbf{L}\mathbf{u}\|^2 \tag{8}$$

Therefore the GCV function to be minimized with respect to μ_{xy} will be given by

$$C_{\mathbf{L}}(\mu_{xy}) = \frac{\|(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H} + \mu_{xy}\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T)\mathbf{b}\|^2}{(\mathrm{tr}(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H} + \mu_{xy}\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T))^2} \tag{9}$$

A classical choice for $\mathbf{L}$ that enforces consistency between the four noisy constraints in (6) would be given by

$$\mathbf{L}^T\mathbf{L} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{10}$$

The main task now is to determine at each pixel the optimal value of μ_{xy} that minimizes the GCV criterion in (9). In existing literature, the minimizer of (9) is usually obtained using numerical techniques. In order to overcome the non-linearity of the GCV criterion, these techniques often make simplifying assumptions, e.g. circulant $\mathbf{H}$ and $\mathbf{L}$, or use numerical techniques such as quadrature rules and Lanczos algorithm [4, 9]. Below, we will derive a closed-form expression for the minimizer of our GCV criterion.

For this purpose, let $\mathbf{P} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H} + \mu_{xy}\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T$. This matrix is commonly referred to as the projection matrix. The GCV function can thus be written as

$$C_{\mathbf{L}}(\mu_{xy}) = \frac{\|\mathbf{P}\mathbf{b}\|^2}{(\mathrm{tr}(\mathbf{P}))^2} \tag{11}$$

Applying the matrix inversion lemma to $\mathbf{P}$, we get

$$\begin{aligned} \mathbf{P} &= \left(\mathbf{I} + \frac{1}{\mu_{xy}}\mathbf{K}\mathbf{K}^T\right)^{-1} \\ &= \mu_{xy}(\mu_{xy}\mathbf{I} + \mathbf{K}\mathbf{K}^T)^{-1} \end{aligned} \tag{12}$$

where $\mathbf{K} = \mathbf{H}\mathbf{L}^{-1}$.

Now, let

$$\mathbf{K}\mathbf{K}^T = \mathbf{V}\Lambda\mathbf{V}^T = \sum_{j=1}^{4} \lambda_j \mathbf{v}_j \mathbf{v}_j^T \tag{13}$$

be the spectral decomposition of $\mathbf{K}\mathbf{K}^T$, where $\mathbf{v}_j$'s are the columns of $\mathbf{V}$ that form a set of orthonormal vectors and λ_j's are the corresponding eigenvalues. It then follows that

$$\mathbf{P} = \sum_{j=1}^{4} \frac{\mu_{xy}}{\mu_{xy} + \lambda_j}\mathbf{v}_j\mathbf{v}_j^T \tag{14}$$

$$= \sum_{j=1}^{2} \frac{\mu_{xy}}{\mu_{xy} + \lambda_j}\mathbf{v}_j\mathbf{v}_j^T + \sum_{j=3}^{4}\mathbf{v}_j\mathbf{v}_j^T \tag{15}$$

where the last equality follows from the fact that in our case (i.e. in eq. (6)) $\lambda_3 = \lambda_4 = 0$.

In order to simplify this GCV function, we take advantage of the fact that most of the information in $\mathbf{K}\mathbf{K}^T$ is in fact carried by the first eigenvalue. In fact, it can be shown that by construction of the matrix $\mathbf{L}$, in the worst case the largest eigenvalue is 3 times the second eigenvalue, but due to lack of space we do not present the formal proof here. Therefore $\mathbf{K}\mathbf{K}^T \simeq \lambda_1\mathbf{v}_1\mathbf{v}_1^T$. Using the abbreviation $\mathbf{V}^T\mathbf{b} = \mathbf{z} = [z_1\ z_2\ z_3\ z_4]^T$, we get

$$C_{\mathbf{L}}(\mu_{xy}) \simeq \frac{\left(\frac{\mu_{xy}}{\mu_{xy}+\lambda_1}\right)^2 z_1^2 + \sum_{j=2}^{4} z_j^2}{\left(3 + \frac{\mu_{xy}}{\mu_{xy}+\lambda_1}\right)^2} \tag{16}$$

Differentiating this equation with respect to μ_{xy} and setting it to zero, we find after simplification that the optimal regularization parameter is given by

$$\mu_{xy}^* = \frac{\lambda_1(z_2^2 + z_3^2 + z_4^2)}{3z_1^2 - z_2^2 - z_3^2 - z_4^2} \tag{17}$$

Therefore, a closed form solution is given by (7) and (17).

3. ERRORS & CONFIDENCE MEASURES

Next, we propose an adaptive scheme to robustify the approach by truncating the quadratic error in (8). To specify an adaptive truncation level at each pixel, we need to determine the pixelwise expected value of error in the estimated motion. Note that our solution at any pixel is of the form

$$\hat{\mathbf{u}} = \mathbf{H}^+\mathbf{b} \tag{18}$$

where the pseudo-inverse $\mathbf{H}^+$ is $(\mathbf{H}^T\mathbf{H} + \mu_{xy}\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T$. We make the following assumptions about the model/data error characterized by e in (6) and the smoothnes of motion field characterized by $\mathbf{a} = \mathbf{Lu}$:

$$\begin{aligned} E[\mathbf{ee}^T] &= \sigma_e^2 \mathbf{I} \\ E[\mathbf{aa}^T] &= \sigma_a^2 \mathbf{I} \\ E[\mathbf{ea}^T] &= 0 \quad \text{(i.e. uncorrelated)} \end{aligned} \tag{19}$$

where $E[\cdot]$ denotes the expected value.

We are interested in the quantity $E[\|\mathbf{u} - \hat{\mathbf{u}}\|^2]$, which with some algebraic manipulations can be found as follows:

$$E\left[\|\mathbf{u} - \hat{\mathbf{u}}\|^2\right] = \\ = \text{tr}(E[\mathbf{uu}^T]) - 2\text{tr}(E[\mathbf{ub}^T]\mathbf{H}^{+T}) + \text{tr}(\mathbf{H}^+ E[\mathbf{bb}^T]\mathbf{H}^{+T}) \tag{20}$$

where $\text{tr}(\cdot)$ stands for the trace of the matrix, and the superscript $+T$ denotes the transpose of the pseudo-inverse.

From the linearity of $E[\cdot]$ it follows that

$$E[\mathbf{uu}^T] = \sigma_a^2(\mathbf{L}^T\mathbf{L})^{-1}, \tag{21}$$

$$E[\mathbf{bb}^T] = \sigma_a^2 \mathbf{H}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T + \sigma_e^2 \mathbf{I}, \tag{22}$$

and

$$E[\mathbf{ub}^T] = \sigma_a^2(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T, \tag{23}$$

where the superscript $-T$ denotes the transpose of the inverse matrix.

Substituting from (21), (22) and (23) into (20), we get

$$E[\|\mathbf{u} - \hat{\mathbf{u}}\|^2] = \sigma_e^2 \text{tr}(\mathbf{H}^+\mathbf{H}^{+T}) + \sigma_a^2 \epsilon^2 \tag{24}$$

where

$$\begin{aligned} \epsilon^2 &= \text{tr}((\mathbf{L}^T\mathbf{L})^{-1}) - 2\text{tr}((\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{H}^{+T}) \\ &+ \text{tr}(\mathbf{H}^+\mathbf{H}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{H}^{+T}) \end{aligned} \tag{25}$$

characterizes the bias in the expected error due to regularization. The bias tends to zero as $\mathbf{H}^+ \rightarrow \mathbf{H}^{-1}$, in which case the expected error will be dominated by the first term characterizing data error. In a "good" scenario, we should be able to trust our data more than our prior information, which implies that

$$\frac{\epsilon^2}{\text{tr}(\mathbf{H}^+\mathbf{H}^{+T})} \leq \mu_{xy}^2 \tag{26}$$

where we have used the fact that $\mu_{xy} = \frac{\sigma_e}{\sigma_a}$.
We use (26) as a measure for evaluating our confidence in a velocity vector. Unreliable, velocity vectors are therefore eliminated by adaptively truncating the quadratic cost function in (8)

$$\hat{\mathbf{u}} = \arg\min Q(\mathbf{u}) \quad \text{if} \quad \frac{\epsilon^2}{\text{tr}(\mathbf{H}^+\mathbf{H}^{+T})} \leq \mu_{xy}^* \tag{27}$$

Note that no knowledge of the actual statistics (i.e. σ_e and σ_a) is required.

4. RESULTS & DISCUSSION

We applied our technique to both synthetic and real data, some of which are shown below. Of course, for the synthetic set since the true motion is known the performance can be quantified and compared to other existing methods. Figures 1 and 2 show the results for the synthetic sequences of Yosemite and the diverging tree, respectively. Tables 1 and 2 compare the average angular error and its standard deviation to those reported by Barron et al. [1] in each case.

Results demonstrate excellent performance for our approach.

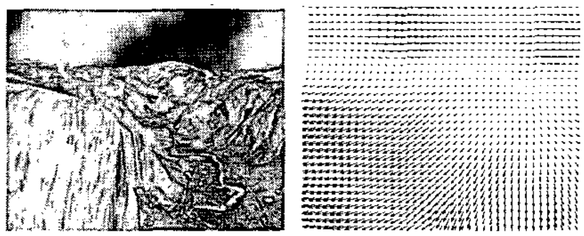


Fig. 1. Yosemite sequence & its estimated flow field.

Technique	Average Error	Standard Deviation
Horn & Schunck	32.43°	30.28°
Lucas & Kanade ($\lambda_2 \geq 1.0$)	4.10°	9.58°
Uras et al. (thresholded)	6.73°	16.01°
Nagel	11.71°	10.59°
Anandan	15.84°	13.46°
Heeger (level 1)	10.51°	12.11°
Fleet & Jepson ($\tau = 2.5$)	4.29°	11.24°
Our Method	0.2°	0.3°

Table 1. Results for Yosemite sequence

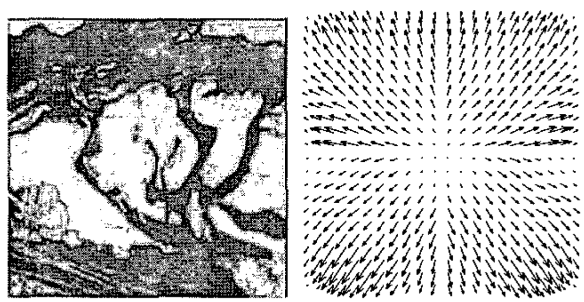
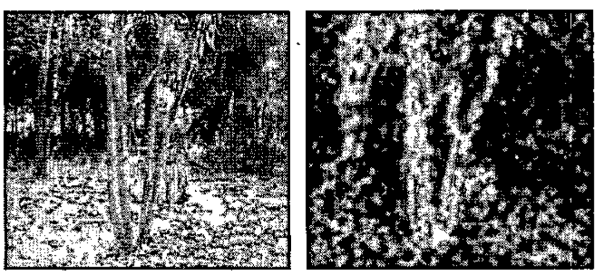


Fig. 2. Diverging tree sequence and its estimated flow field.

Technique	Average Error	Standard Deviation
Horn & Schunck	12.02°	11.72°
Lucas & Kanade ($\lambda_2 \geq 1.0$)	1.94°	2.06°
Uras et al. (thresholded)	3.83°	2.19°
Nagel	2.94°	3.23°
Anandan	7.64°	4.96°
Heeger	4.95°	3.09°
Fleet & Jepson ($\tau = 2.5$)	0.99°	0.78°
Our Method	0.1°	0.25°

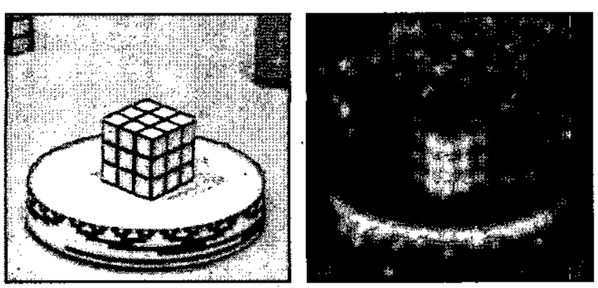
Table 2. Results for diverging tree sequence

We show our results on two real sequences. The first sequence is the SRI tree sequence which has been used by most authors in the past for testing their algorithms in terms of preserving discontinuities and avoiding over-smoothing. The second real sequence is the Rubic cube sequence which has a rotational motion on a turn-table. In both cases we have shown the magnitude of the estimated flow fields to illustrate the capability of our technique for preserving fine scene-related features such as discontinuities.



(a) (b)

Fig. 3. (a) SRI tree sequence, (b) the magnitude of the flow field in gray scales



(a) (b)

Fig. 4. (a) Rubic cube sequence, (b) the magnitude of the flow field in gray scales

For instance in the SRI tree sequence one can easily distinguish in the magnitude image the tree in the middle and its branches and its leaves. In addition the second tree to the left and also an object to the right of the middle tree are also easily distinguished. Figure 4 shows the results for the Rubic cube sequence. One can again easily recognize the patterns on the cube reflected in the magnitude image, which is due to adaptive regularization.

We found that Using GCV and the fact that second order constraints (together with first order constraints) provide pixelwise formulation, one obtains a very attractive framework for solving the problem. In particular such framework alleviates the artificially introduced issues due to neighborhood formulation, while providing a truely pixelwise adaptive solution.

5. REFERENCES

[1] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Int. J. Computer Vision*, 12(1):43–77, 1994.

[2] M.J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. Computer Vision*, pages 231–236, 1993.

[3] H. Foroosh. A closed-form solution for optical flow by imposing temporal constraints. In *Proc. ICIP*, pages 656–659, 2001.

[4] G.H. Golub and M. Heath. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):pp. 195–197, 1979.

[5] D.J. Heeger. Model for the extraction of image flow. *J. Opt. Soc. Am. A*, 4:pp. 1455–1471, 1987.

[6] B.K.P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, vol. 17:pp. 185–203, 1981.

[7] H.H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:pp. 299–324, 1987.

[8] S. Negahdaripour. Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. *IEEE. Trans. Pattern Analysis Machine Intelligence*, 20(9):961–979, 1998.

[9] N. Nguyen, P. Milanfar, and G.H. Golub. Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. *IEEE Trans. Image Proc.*, 10(9):1299–1308, 2001.

[10] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biol. Cybern.*, 60:79–97, 1988.