Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 53912, 9 pages
doi:10.1155/2007/53912

Research Article

Better Flow Estimation from Color Images

Hui Ji[1] and Cornelia Fermüller[2]

[1] *Department of Mathematics, National University of Singapore, Singapore 117543*
[2] *Computer Vision Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742-3275, USA*

Received 1 October 2006; Accepted 20 March 2007

Recommended by Nicola Mastronardi

One of the difficulties in estimating optical flow is bias. Correcting the bias using the classical techniques is very difficult. The reason is that knowledge of the error statistics is required, which usually cannot be obtained because of lack of data. In this paper, we present an approach which utilizes color information. Color images do not provide more geometric information than monochromatic images to the estimation of optic flow. They do, however, contain additional statistical information. By utilizing the technique of instrumental variables, bias from multiple noise sources can be robustly corrected without computing the parameters of the noise distribution. Experiments on synthesized and real data demonstrate the efficiency of the algorithm.

1. INTRODUCTION

Optical flow estimation is a heavily studied problem in computer vision. It is well known that the problem is difficult because of the discontinuities in the scene. However, even at the locations of smooth scene patches, the flow cannot be estimated very accurately because of statistical difficulties.

In this paper, we consider gradient-based approaches to optical flow estimation. The estimation is based on the basic constraint of constant brightness at an image point over a small time interval. This can be expressed as follows [1]:

$$I_x u_x + I_y u_y + I_t = 0, \tag{1}$$

where I_x, I_y, I_t denote the spatial and temporal derivatives of the image intensity function I, and $\vec{u} = (u_x, u_y)$ denotes the velocity vector at an image point. This equation, known as the *brightness consistency constraint*, only gives one component of the optical flow vector $\vec{u} = (u_x, u_y)$. To obtain the second component, further assumptions on the optical flow need to be imposed. Common nonparametric constraints are obtained by assuming that the flow field is smooth locally (see [2] for a comprehensive survey). Other approaches assume a parametric model for the optical flow. Regardless of the strategy adopted, one usually arrives at an overdetermined linear equation system of the form

$$A\vec{x} = \vec{b}, \tag{2}$$

where $\vec{x}$ denotes the parameter vector characterizing the optical flow, and A and $\vec{b}$ are the measurements. For example, for the model of constant flow in a spatial neighborhood, assuming we combine n image brightness constraint equations, A becomes the $n \times 2$ matrix (I_{x_i}, I_{y_i}), $\vec{b}$ is the n-dimensional vector $(-I_{t_i})$, and $\vec{x}$ is just the two-dimensional optical flow $\vec{u} = (u_x, u_y)$. If the model assumes the flow to be a polynomial function in the image coordinates, then the flow components u_x and u_y are linear with respect to some $k \times 1$ parameter vector $\vec{x}$. In other words,

$$\begin{aligned} u_x &= \vec{f_i}^t \vec{x}, \\ u_y &= \vec{g_i}^t \vec{x}, \end{aligned} \tag{3}$$

where $\vec{f_i}$, $\vec{g_i}$ are $k \times 1$ vectors of image coordinates. For example, if the scene patch in view is a plane, the flow model amounts to

$$\begin{aligned} \vec{f_i} &= \left(x_i, y_i, -1, 0, 0, 0, x_i^2, x_i y_i\right)^t, \\ \vec{g_i} &= \left(0, 0, 0, x_i, y_i, -1, x_i y_i, y_i^2\right)^t. \end{aligned} \tag{4}$$

A is composed of the spatial image gradients and image coordinates, $\vec{b}$ still encodes the temporal derivatives, and $\vec{x}$ are the coefficients of the polynomial.

The most common approach to estimating the parameter vector $\vec{x}$ is by means of least squares (LS) regression. However, LS implicitly makes the assumption that the explanatory variables, that is the elements of A in (1), are measured without errors. This is not true, because the spatial derivatives I_x, I_y are always noisy. Thus, we are dealing with what is called the *errors-in-variables* (EIV) noise model in Statistics. LS estimation on this model can be shown to be inconsistent, and the bias leads to underestimation.

The bias in the LS estimation of optical flow has been observed before (see [3, 4]), and alternative estimators have been proposed. In particular, total least squares estimation (TLS) has received significant attention. A straightforward approach of TLS is problematic, as TLS assumes the noise components in the spatial and temporal derivatives to be pixel-wise independent. References [3, 5] addressed the correlation of the image derivatives between pixels using a maximum-likelihood (MLE) estimator. Reference [6] developed the so-called heteroscedastic errors-in-variable (HEIV) estimator. In essence, both approaches are modifications of TLS estimation to account for the underlying noise processes with pixel-wise dependence and nonhomogeneous covariance. The tradeoff for dealing with these dependences is higher complexity and less stability in the resulting procedures. Furthermore, the corresponding objective functions are nonlinear and nonconvex, which makes the minimization a difficult task.

Most studies of optical flow utilize gray-scale image sequences, but color image sequences have been used as well (see [7, 8]). A common approach to utilizing the color is to incorporate more constraints into the optical flow computation. Essentially, one color sequence provides three image sequences. Another approach is to substitute the brightness consistency constraint by a color consistency constraint to obtain equations with higher accuracy. However, previous studies did not consider noise in the color images, or extracting statistical information from color.

Three color channels do not contain more information than one mono-chromatic channel from a geometric point of view. They do, however, contain statistical information. Here we use this color information to correct for the bias in optical flow. The approach is based on the so-called instrumental variable (IV) estimator, which has several advantages over other estimators. Most important, it does not require an estimation of the error, and it can handle multiple heteroscedastic noise terms. Furthermore, its computational complexity is comparable to LS.

After giving a brief introduction to the EIV model and LS and TLS regression in Section 2, we discuss the sources of noise in Section 3. Then in Section 4 we present our IV method on color image sequences. Section 5 compares the performance of our IV method against LS and TLS estimation, and Section 6 concludes the paper.

2. THE ERROR MODEL AND COMMON ESTIMATORS

The problem of estimating optical flow from the brightness consistency constraints amounts to finding the "best" solution to an over-determined equation system of the form $A\vec{x} = \vec{b}$. The observations A and $\vec{b}$ are always corrupted by errors, and in addition there is system error. We are dealing with what is called the errors-in-variable (EIV) model in statistical regression, which is defined as follows [9].

Definition 1 (error-in-variables model).

$$\begin{aligned} \vec{b'} &= A'\vec{x'} + \vec{\epsilon}, \quad \text{with } \vec{b'} \in \mathbb{R}^n,\ A' \in \mathbb{R}^{n\times k}, \\ \vec{b} &= \vec{b'} + \vec{\delta_b}, \\ A &= A' + \delta_A. \end{aligned} \tag{5}$$

$\vec{x'}$ are the true but unknown parameters. $A = (A_{i,j})$ and $\vec{b} = (b_i)$ are observations of the true but unknown values A' and $\vec{b'}$. $\delta_A = (\delta_{A_{i,j}})$ and $\vec{\delta_b} = (\delta_{b_j})$ are the measurement errors and $\vec{\epsilon} = (\epsilon_i)$ is the system error, also called equation error or modeling error. This is the error due to the model assumptions.

In the following discussion, we consider $\delta_{A_{i,j}}$, δ_{b_i}, and ϵ_i to be independent and identically distributed random variables with zero mean and variances σ_A^2, σ_b^2, and σ_ϵ^2, respectively.

The most popular choice to solving the system is by means of least squares (LS) estimation which is defined as

$$\vec{x}_{\text{LS}} = (A^tA)^{-1}(A^t\vec{b}). \tag{6}$$

However, it is well known that the LS estimator $\vec{x}_{\text{LS}}$ is generally biased [10]. Consider the simple case where all elements in δ_A and $\vec{\delta_b}$ are independent and identically distributed random variables with zero mean and variance σ^2. Then we have

$$\lim_{n\to\infty} E(\vec{x}_{\text{LS}} - \vec{x'}) = -\sigma^2\left(\lim_{n\to\infty}\left(\frac{1}{n}A'^tA'\right)\right)^{-1}\vec{x'}, \tag{7}$$

which implies that $\vec{x}_{\text{LS}}$ is asymptotically biased. Large variance in δ_A, an ill-conditioned A, or an $\vec{x}$ which is oriented close to the eigenvector of the smallest singular value of A all could increase the bias and push the LS solution $\vec{x}_{\text{LS}}$ away from the real solution. Generally, it leads to an underestimation of the parameters.

The so-called corrected least squares (CLS) estimator theoretically could correct the bias. Assume σ, the variance of the error, to be known a priori. Then the CLS estimator for $\vec{x}$, which is defined as

$$\vec{x}_{\text{CLS}} = (A^tA - n\sigma^2 I)^{-1}(A^t\vec{b}), \tag{8}$$

gives asymptotically unbiased estimation. This estimator is also known as *correction for attenuation* in Statistics. The problem is that accurate estimation of the variance of the error is a challenging task. Since the scale of the error variance is difficult to obtain in practice, this estimator is not very popular in computer vision.

Since the exact error variance is hard to obtain, the so-called total least squares (TLS) or orthogonal least squares estimator, which only requires the estimation of the ratio $\eta = (\sigma_b^2 + \sigma_\epsilon^2)/(\sigma_A^2)$, became popular. It can be formulated as the following nonlinear minimization:

$$\vec{x}_{\text{TLS}} = \arg_{\vec{x}} \min M(\vec{x}, \eta) = \arg_{\vec{x}} \min \sum_i \frac{1}{n} \frac{||A_i\vec{x} - b_i||^2}{||\vec{x}||^2 + \eta}. \tag{9}$$

The vector $(\eta^{-1/2}\vec{x}_{\text{TLS}}; -1)$ is the singular vector corresponding to the smallest singular value of the SVD of the matrix $[A, \eta^{-1/2}b]$. If the errors $\delta_A, \vec{\delta_b}$ are independent and identically distributed, or if we can obtain the ratio $\eta = (\sigma_b^2 + \sigma_\epsilon^2)/\sigma_A^2$, then TLS estimation is asymptotically unbiased. However, the main problem for TLS is system error, that is $\vec{\epsilon}$. One may simply merge the equation error ϵ_i with the measurement error b_i into a new error variable, as it is done in least squares estimation. But this affects the estimation. System error is due to the fact that our model is only some approximation of the underlying real model. We can have multiple tests to obtain the measurement error, like remeasuring or resampling; but unless we know the exact parameters of the model, we cannot obtain the system error. If the equation error is simply omitted, the estimation becomes an *overestimation* (see [11]). Thus, unless the system error is small and accurate estimation of the ratio of variances can be obtained accurately, TLS will not be unbiased.

Another problem with TLS for computer vision applications is that often the noise is *heteroscedastic* [6]. In other words, the noise is independent for each variable, but correlated for the measurements. Although we still could apply TLS (assuming we normalize for the different variances in the noise), the corresponding objective function is nonlinear and nonconvex. As shown in [12], the *long valley* in the objective function surface around the minimum point often causes a problem in the convergence. If, however, the error is mismodeled, the performance of TLS can decrease very much.

3. NOISE

Now let us investigate a realistic error model for our flow equation

$$A\vec{x} = \vec{b}. \tag{10}$$

This equation is based on two assumptions:

(1) *intensity consistency*: the intensity of a point in the image is constant over some time interval,
(2) *motion consistency*: the motion follows some model. For example, the flow is approximated by a polynomial function in the image coordinates, or the flow varies smoothly in space.

The errors, thus, can be categorized into

(1) *modeling error*: the intensity is not constant or the motion model fails to fit the real motion,
(2) *measurement noise*: this is mainly sensor noise and noise due to poor discrete approximation of the image derivatives I_x, I_y, and I_t.

We argue that we need to take both kinds of error into account. Modeling errors always occur. They are associated with the scene and its geometrical properties. Modeling errors become large at specularities and at the boundaries between two different regions, or if the model does not apply. These errors have much less randomness than the measurement noise. The measurement noise generally can be treated as random variables. Most studies only consider measurement noise, and only the one due to sensors (see [6, 12]). But we want to deal with all the sources of noise. In general we are facing a combination of multiple heteroscedastic noises. We could attempt to use a sophisticated noise model. But it appears too complicated to estimate the variances of the different noise components, which is necessary to apply CLS or TLS regression. Fortunately, we do not need to. In the next section we will introduce a parameter regression called the *instrumental variables method* (IV), which has been used extensively in economics.

4. COLOR IMAGES AND IV REGRESSION

4.1. IV regression

As regression model we have the EIV model as defined in Definition 1, with $A \in \mathbb{R}^{n\times k}$, $\vec{b} \in \mathbb{R}^n$, and $\vec{x'} \in \mathbb{R}^k$.

Definition 2 (instrumental variables method). Consider a random sample matrix W of $n\times j$ elements, called the *instruments* or *instrumental variables* of A, which has the following properties:

(1) $E(W^t(\delta_A\vec{x} + \vec{\delta_b})) = 0$.
(2) $E((W^tW)^{-1}W^tA')$ is a nonsingular matrix.

Then the instrumental variable estimator for the k dimensional vector $\vec{x}$, defined as

$$\vec{x} = \begin{cases} (W^tA)^{-1}W^t\vec{b} & \text{if } j = k, \\ [A^tW(W^tW)^{-1}W^tA]^{-1} & \\ \quad A^tW(W^tW)^{-1}W^t\vec{b} & \text{if } j \geq k \end{cases} \tag{11}$$

is an asymptotically unbiased estimator for $\vec{x'}$. The variance of the estimated $\vec{x}$ can be estimated as

$$V(\vec{x}) = \frac{1}{n-k}(\hat{A}^t\hat{A})^{-1}\sum_{i=1}^{n}(\vec{b_i} - A_i^t\vec{x})^2, \quad \text{with } \hat{A} = W(W^tW)^{-1}W^tA, \tag{12}$$

when $j = k$.

Let us explain this model. Intuitively, two things are required for a set of measurements W to be instrumental variables to the original measurements. The first one is that the instrumental variables are not correlated with the noise terms in the estimation model. The second one is that the instrumental variables and the explanatory variables are not independent, and thus the correlation matrix has full rank, and that W has full column rank. Then instead of premultiplying, as in LS, (2) with A^t to derive at $A^tA\vec{x} = A^t\vec{b}$, we premultiply with W^t to solve an equation system of the form

$$(W^tA)\vec{x} = W^t\vec{b}. \tag{13}$$

This is easy if the number of variables in A and W is the same. For example, if A and W result from two ways of estimating A'. But the number of variables could also be different. In this case, most often the IV method is implemented as a two-stage regression. In the first stage A and $\vec{b}$ are regressed on the instrumental variables. Requirement 2 guarantees that the instrumental variables are related to A and $\vec{b}$. In the second stage the regression of interest is estimated as usual, except that now each covariate is replaced with its approximation estimated in the first stage. Requirement 1 guarantees that the noise in this stage does not make the estimation biased. More clearly, rewrite the regression as a new regression model:

$$\begin{aligned} \vec{b} &= W\vec{\pi}_1, \\ A &= W\Pi_2. \end{aligned} \tag{14}$$

Then the first regression yields

$$\begin{aligned} \vec{\pi}_1 &= (W^tW)^{-1}W^t\vec{b}, \\ \Pi_2 &= (W^tW)^{-1}W^tA. \end{aligned} \tag{15}$$

The relation between $\vec{\pi}_1$ and Π_2 is $\vec{\pi}_1 = \Pi_2\beta$. Then the least squares estimator in the second stage gives

$$\vec{x} = (\Pi_2^t\Pi_2)^{-1}\Pi_2^t\vec{\pi}_1. \tag{16}$$

Mathematically this estimator is identical to a single stage estimator when the number of instruments is the same as the number of covariates (i.e., $j = k$), leading to the formulation of (11).

The technique of instrumental variables is highly robust to improper error modeling. It can be used even if the instrumental variables are not completely independent of the measurement errors. The worst that can happen is that A and W have the exact same measurement error, in which case the method reduces to LS estimation. To summarize, the advantages of IV regression over other techniques are the following.

(1) It does not require assumptions about the distribution of the noise.
(2) It can handle multiple heteroscedastic noise terms. In comparison, other methods need to derive specific complicated minimization procedures for the specific problem.
(3) The minimization is simple and noniterative with a computational complexity which is comparable to LS.

Next we show how to construct appropriate instrumental variables for the estimation of the optic flow parameters.

4.2. Color images

Here we consider an RGB color model. Other color models are similar. The RGB model decomposes colors into their red, green, and blue components (R,G,B). Thus, from the brightness consistency constraint we can obtain three linear equation systems:

$$\begin{aligned} A_R\vec{x} &= \vec{b}_R, \\ A_G\vec{x} &= \vec{b}_G, \\ A_B\vec{x} &= \vec{b}_B. \end{aligned} \tag{17}$$

Why should the A_R, A_G, A_B be appropriate instrumental variables to each other? For a natural scene, the correlation between the image gradients of the three color images is very high. Therefore the second requirement for instrumental variables is satisfied in most cases. And what about the first requirement, that is, the independence of the noise terms? It is quite reasonable to assume that the sensor noise components are independent if the sequence is taken by a true color camera. The approximation errors in the image gradients will not be completely independent, since there is a similarity in the structure of the color intensity functions. We found in our experiments, that for scenes with noticeable color variation, the correlation between the approximation errors is rather weak. This means that we cannot completely remove the bias from approximation error, but we can partially correct the bias caused by this error. We cannot correct the bias from the modelling error. But despite the presence of modelling error, we still can deal with the other errors. Other estimators like TLS cannot.

Using the image gradients of one color channel as instrumental variables to the image gradients of another color channel we obtain six different IV estimations of the real value $\vec{x}'$:

$$\begin{aligned} \vec{x}_1 &= (A_B^tA_R)^{-1}(A_B^t\vec{b}_R), & \vec{x}_2 &= (A_G^tA_R)^{-1}(A_G^t\vec{b}_R), \\ \vec{x}_3 &= (A_G^tA_B)^{-1}(A_G^t\vec{b}_B), & \vec{x}_4 &= (A_R^tA_B)^{-1}(A_R^t\vec{b}_B), \\ \vec{x}_5 &= (A_B^tA_G)^{-1}(A_B^t\vec{b}_G), & \vec{x}_6 &= (A_R^tA_G)^{-1}(A_R^t\vec{b}_G). \end{aligned} \tag{18}$$

Because of small sample size, in practice we use Fuller's modified IV estimator [9], which is defined as

$$\vec{x} = [\hat{A}^t\hat{A} - \nu S_{22}]^{-1}[\hat{A}^t\hat{b} - \nu S_{21}], \tag{19}$$

where ν is a constant term that we set equal to 1. Our experiments found that the choice of ν does not have a significant influence on the estimation. $\hat{A}$, $\hat{\vec{b}}$ are defined as

$$(\hat{A}, \hat{\vec{b}}) = W(W^tW)^{-1}W^t(A, \vec{b}) \tag{20}$$

and $S_{21} \in \mathbb{R}^{2\times 1}$ and $S_{22} \in \mathbb{R}^{2\times 2}$ are submatrices of

$$S = (n-k)^{-1}[(\vec{b}, A)^t(\vec{b}, A) - (\vec{b}, A)^tW(W^tW)^{-1}W^t(\vec{b}, A)]. \tag{21}$$

Now we have six estimations for $\vec{b}$, or even nine if we include the three least squares estimates. We can also estimate their respective variances $V(\vec{x})$ according to (12), and take the weighted mean of these estimates as our final estimate:

$$\vec{x} = \left(\sum_{i=1}^{6} V(\vec{x}_k)^{-1}\right)^{-1} \sum_{k=1}^{6} V(\vec{x}_k)^{-1}\vec{x}_k. \tag{22}$$

Such an estimate $\vec{x}$ will decrease the variance and effectively correct the bias.

4.3. *Robust IV regression*

So far, we have only discussed small-scale noise. Often, we also have large-scale measurement errors (outliers). Such errors occur in the temporal derivatives at the motion boundaries or in the spatial derivatives close to the boundary of objects. Outliers will seriously decrease the performance of any estimator, LS, TLS, as well as the IV estimator. Next, we discuss an IV version of robust regression.

A popular robust regression dealing with outliers is the median regression. Assuming that δA_i, $\vec{\delta}_{b_i}$, and W_i are independent, we obtain that

$$\mathrm{Med}([\delta A_i, \vec{b}_i] \mid W_i) = 0, \tag{23}$$

which implies that

$$E[\,\mathrm{sgn}\,(W_i^t(\vec{b}_i - A_i^t\vec{x}))] = 0. \tag{24}$$

Then a robust estimator $\vec{x}_{M\,\mathrm{IV}}$ can be defined as the minimum of some norm of its sample analogue:

$$\frac{1}{n}\sum_{i=1}^{n} W_i^t(\mathbf{1}\{\vec{b}_i - A_i^t\vec{x} > 0\} - \mathbf{1}\{\vec{b}_i - A_i^t\vec{x} < 0\}), \tag{25}$$

where $\mathbf{1}\{\cdot\}$ is defined as

$$\mathbf{1}\{\Gamma\} = \begin{cases} 1, & \Gamma \text{ is TRUE}, \\ 0, & \text{Otherwise}. \end{cases} \tag{26}$$

Such a robust estimator $\vec{x}_{M\,\mathrm{IV}}$ effectively detects outliers. After eliminating the outliers, the usual IV estimation can be applied to obtain an accurate estimation of the parameters.

4.4. *Integration of IV regression into differential flow algorithms*

A very popular optical flow model is the weighted local constant flow model, where one minimizes

$$\sum_i w_i^2(\nabla I_i^t\vec{x} + I_{t_i})^2 \tag{27}$$

with $\nabla I_i = (I_{x_i}, I_{y_i})$ denoting the spatial derivatives of I. It is easy to see that this amounts to the usual least squares regression for the linear system $A\vec{x} = \vec{b}$, with

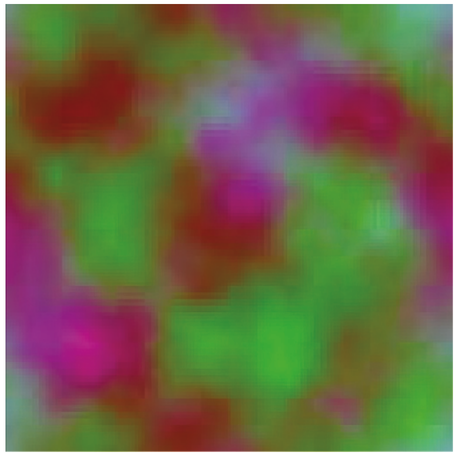


Figure 1: Reference images for the "cloud" sequence.

$$\begin{aligned} A &= (w_i\nabla I_i), \\ \vec{b} &= -(w_i I_{t_i}), \\ \vec{x} &= (u_x, u_y). \end{aligned} \tag{28}$$

We can apply the IV regression to any combination of two colors. For example, we can take color channels R and B to obtain

$$A_{\mathrm{R}}^t A_{\mathrm{B}}\vec{x} = A_{\mathrm{R}}^t\vec{b}_{\mathrm{B}}, \tag{29}$$

with

$$\begin{aligned} A_{\mathrm{R}} &= (w_{i_1}\nabla \mathrm{R}_i), \\ A_{\mathrm{B}} &= (w_{i_2}\nabla \mathrm{B}_i), \\ \vec{x}_{\mathrm{R}} &= (w_{i_1}\mathrm{R}_{t_i}), \\ \vec{x}_{\mathrm{B}} &= (w_{i_2}\mathrm{B}_{t_i}). \end{aligned} \tag{30}$$

Another common model assumes the surface in view to be a parametric function of the image coordinates. For example, if the surface is fronto parallel, the flow is linear. If the surface is a slanted plane, the flow is quadratic. Such flow models often are used in image registration and egomotion estimation. The corresponding brightness consistency constraint $\nabla I^t\vec{x} + I_t = 0$ still is a linear system with a parameter vector that encodes motion and surface information.

We also could easily incorporate the IV regression into flow algorithms which enforce some smoothness constraints. We only need to replace the LS form for the brightness consistency constraint by its IV form while leaving the smoothness penalty part of the objective function in the minimization the same.

5. EXPERIMENTAL RESULTS AND SUMMARY

We compared the performance of IV estimation against LS estimation and a straight forward version of TLS estimation with similar complexity.

Using the two images in Figures 1 and 2, we generated image sequences with 2D rigid motion, that is, 2D rotation

Figure 2: Reference image for the "office" sequence.

and translation, that is, the image motion amounts to

$$\begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \tag{31}$$

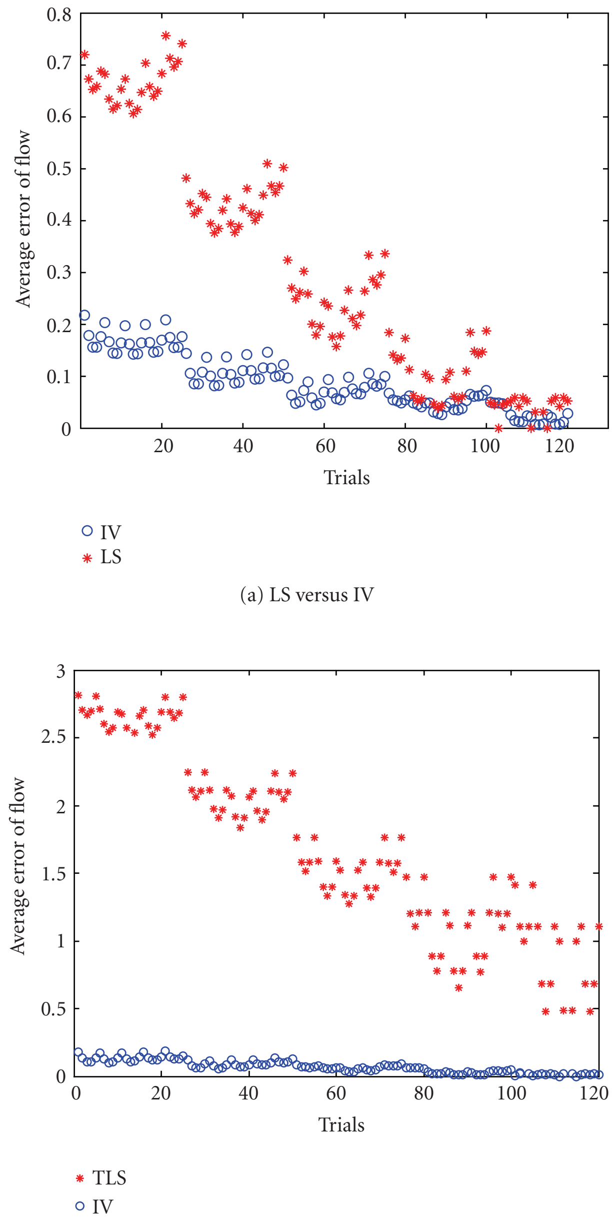
In the first experiment we described the flow with an affine model as

$$\begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}. \tag{32}$$

Thus, having computed the image derivatives I_x, I_y, I_t, the estimation amounts to first finding the parameters (a, b, t_x, t_y) and then computing the optical flow at every point from (32). The average error is defined as the average difference between the estimated optical flow and the ground truth (over all pixels). In total 150 motion sequences were created with parameter α distributed uniformly between $[-5, 0]$, and parameters t_x, t_y distributed uniformly between $[-1, 1]$.

The scatter plots of the average error in Figure 3 clearly show the advantage of IV over LS and TLS. The performance of TLS is much worse than LS, which from our discussion in the previous section, is not surprising. (The normalization is critical for the success of TLS. However, it also increases the complexity dramatically.) The improvement of IV over LS in Figure 4 is not as good as in Figure 3. This is due to the fact that the three color channels in the sequence "office" (see Figure 2) in many locations are very similar to each other, while the three color channels in the sequence "cloud" (see Figure 1) are not. Thus, the overall effect of bias correction is less. But the IV method still could achieve moderate bias correction.

In the second experiment, we used the Lucas-Kanade multiscale algorithm [13], which does not rely on a parametric predefined flow model. We used a pyramid with three levels of resolution and we added Gaussian noise with $\sigma = 4$ to the synthetic image sequences. (This corresponds to an SNR defined as I_{mean}/σ of $(25.7, 23.9, 14.7)$ for the three channels (R,G,B), or a PSNR (peak SNR) defined as $20\log_{10}(255/\sigma^2)$ of 24 for all three color channels.) 54 trials were conducted



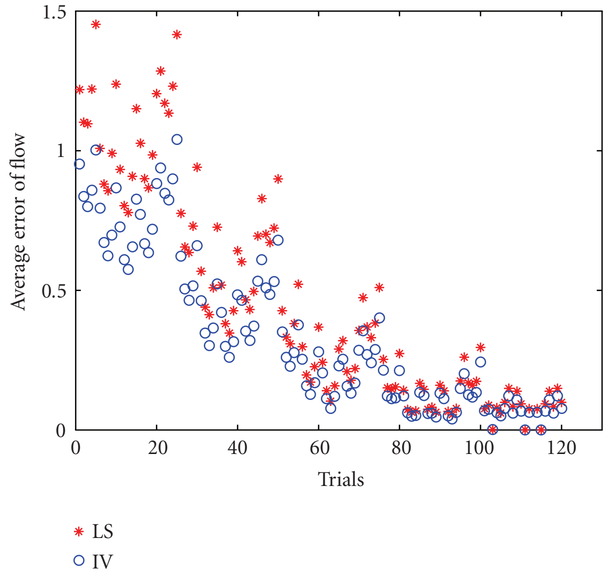
(a) LS versus IV

(b) TLS versus IV

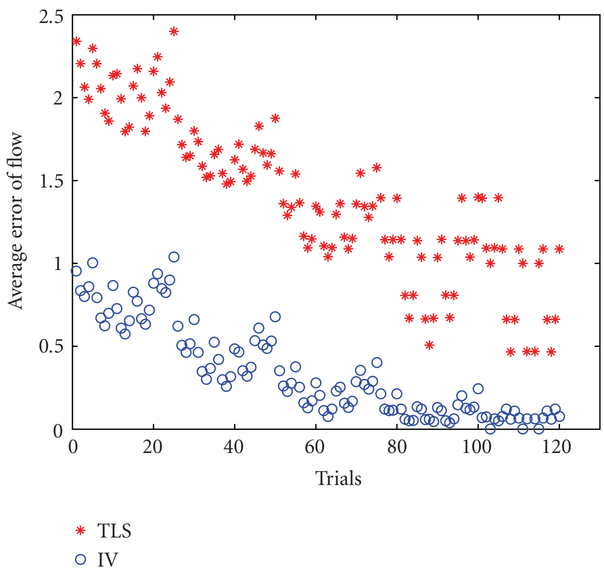
Figure 3: Performance comparison on the "cloud" sequence.

with randomly chosen 2D motion parameters in the same intervals as in the first experiment. The average errors in the optical flow are illustrated in Figures 5 and 6. As in the first experiment, the IV method outperforms the other two methods, and the improvement is much larger for the "cloud" sequence than for the "office" sequence.

We also compared the three flow estimators on a real image sequence. A robot moved with controlled translation in the corridor carrying a camera that pointed at some angle at a wall, which was covered with magazine paper (see Figure 7 for one frame). The camera was calibrated, and thus the ground truth of the optical flow was known. The flow was estimated using the Lucas-Kanade multiscale algorithm with

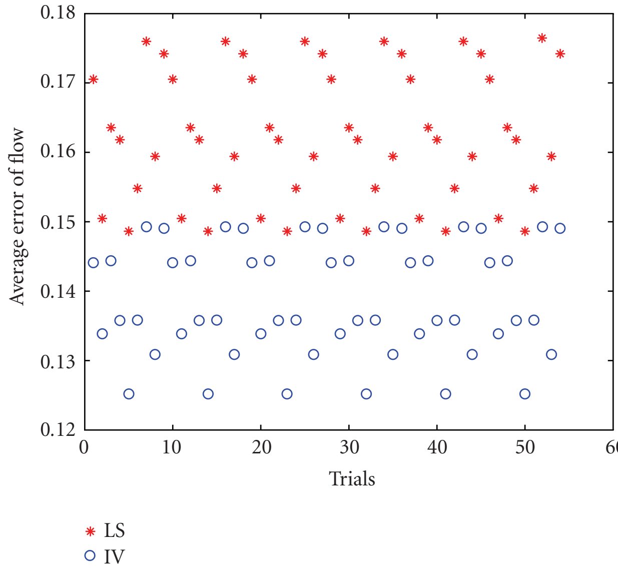


(a) LS versus IV

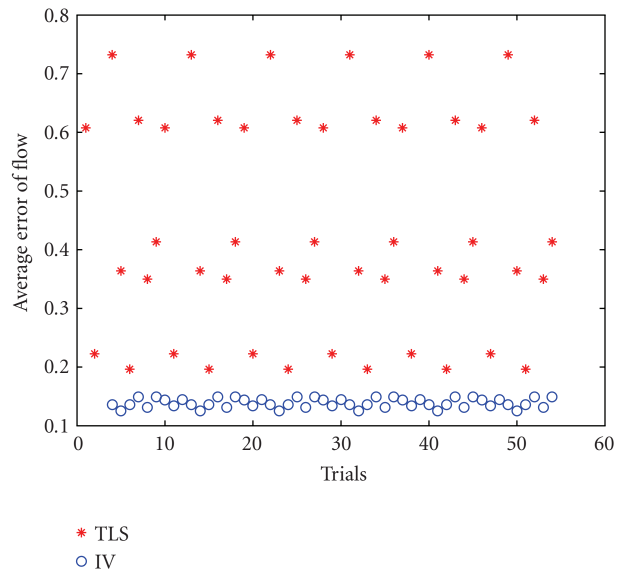


(b) TLS versus IV

FIGURE 4: Performance comparison on the "office" sequence.



(a) LS versus IV



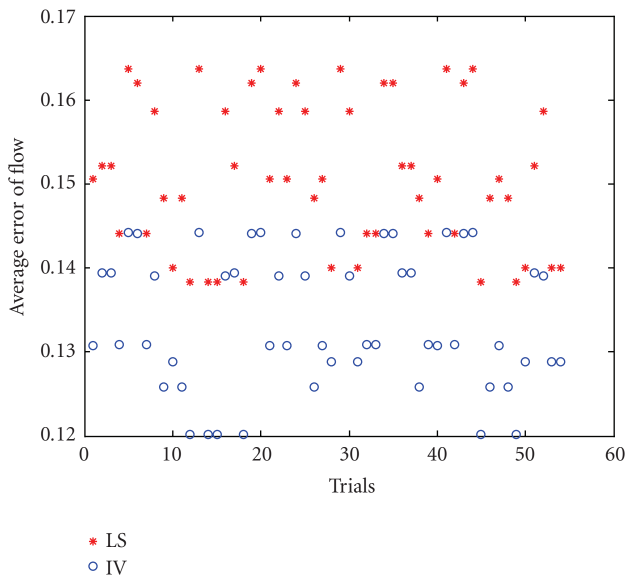
(b) TLS versus IV

FIGURE 5: Performance comparison on the "cloud" sequence using the Lucas-Kanade algorithm.

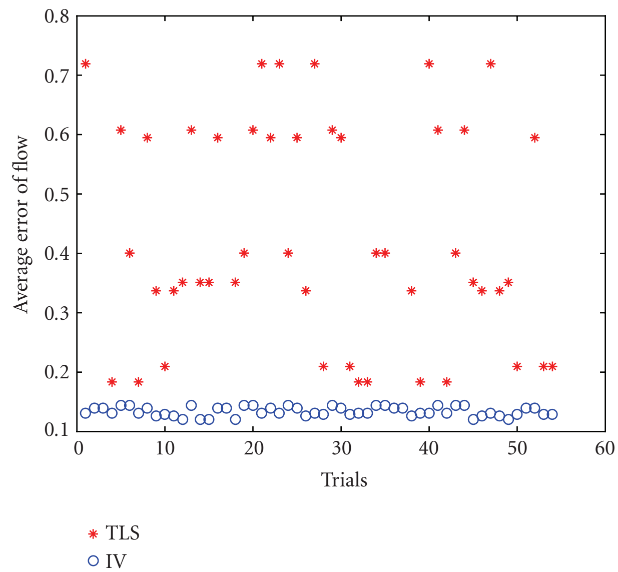
three levels of resolution. The estimation was performed on the individual color channels (R, G, B) and on the combined color channels using LS, TLS, and the IV method. Figure 8(a) shows the average angular error between the estimated flow and the ground truth. Figure 8(b) shows the average relative error in the magnitude of the horizontal flow component, that is, denoting the ground truth of the magnitude of the flow values as x_i', the estimates as x_i, and the confidence in the estimates (the condition number of matrix A in the equation $Ax = b$) as con (x_i), the error was found as the mean of $(|x' - x_i|/x_i)$con (x_i). As can be seen, there is different information in the individual color channels. However, how to fuse the three channels, to arrive at more accurate estimates, is not a trivial task. The IV method performed best among the three estimators in fusing the color channels.

6. CONCLUSIONS

We presented a new approach to correct the bias in the estimation of optical flow by utilizing color image sequences. The approach is based on the instrumental variables technique. It is as simple and fast as ordinary LS estimation, while providing better performance. The same technique could also be applied to other estimation problems in image reconstruction. For example, the estimation of shape from different cues, such as stereo, texture, or shading. Many of the



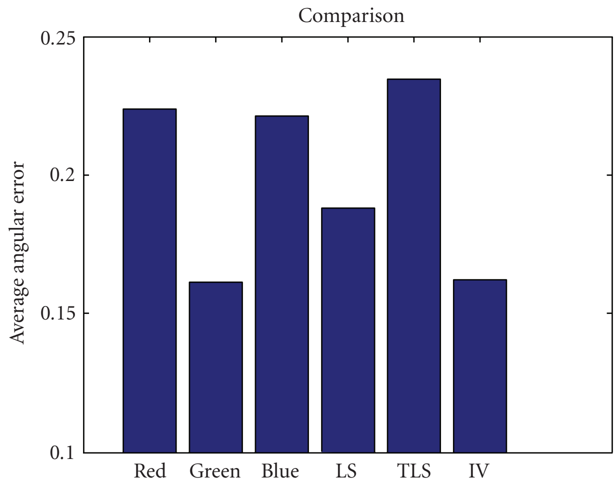
(a) LS versus IV



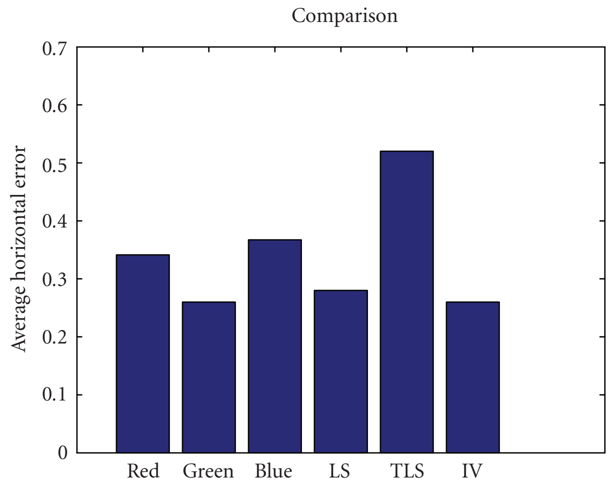
(b) TLS versus IV

Figure 6: Performance comparison on the "office" sequence using the Lucas-Kanade algorithm.

Figure 7: One frame in the "wall" sequence.



(a)



(b)

Figure 8: Performance comparison on the "wall" sequence: (a) Average angular error (in degrees) between estimation and ground truth. (b) Average relative error in value of horizontal flow component.

shapes from X techniques employ linear estimations, or they use regularization approaches, which also could incorporate a bias correction in the minimization.

REFERENCES

[1] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

[2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[3] H.-H. Nagel, "Optical flow estimation and the interaction between measurement errors at adjacent pixel positions," *International Journal of Computer Vision*, vol. 15, no. 3, pp. 271–288, 1995.

[4] C. Fermüller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow," *Computer Vision and Image Understanding*, vol. 82, no. 1, pp. 1–32, 2001.

[5] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Oxford, UK, 1996.
[6] J. Bride and P. Meer, "Registration via direct methods: a statistical approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 984–989, Kauai, Hawaii, USA, December 2001.
[7] R. J. Andrews and B. C. Lovell, "Color optical flow," in *Proceedings of the Workshop on Digital Image Computing*, vol. 1, pp. 135–139, Brisbane, Australia, February 2003.
[8] P. Golland and A. M. Bruckstein, "Motion from Color," *Computer Vision and Image Understanding*, vol. 68, no. 3, pp. 346–362, 1997.
[9] W. A. Fuller, *Measurement Error Models*, John Wiley & Sons, New York, NY, USA, 1987.
[10] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, vol. 9 of *Frontiers in Applied Mathematics Series*, SIAM, Philadelphia, Pa, USA, 1991.
[11] R. J. Carroll and D. Ruppert, "The use and misuse of orthogonal regression estimation in linear errors-in-variables models," Tech. Rep., Department of Statistics, University of Texas A&M, College Station, Tex, USA, 1994.
[12] L. Ng and V. Solo, "Errors-in-variables modeling in optical flow estimation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1528–1540, 2001.
[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pp. 674–679, Vancouver, BC, Canada, August 1981.

Hui Ji received his B.S. degree, M.S. degree in mathematics and Ph.D. degree in computer science from Nanjing University, National University of Singapore, and the University of Maryland at College Park, respectively. Since 2006 he has been an Assistant Professor in the Department of Mathematics at the National University of Singapore. His research interests are in human and computer vision, image processing, and computational harmonic analysis.

Cornelia Fermüller received the M.S. degree in applied mathematics from the University of Technology, Graz, Austria in 1989 and the Ph.D. degree in computer science from the Technical University of Vienna, Austria in 1993. Since 1994 she has been with the Computer Vision Laboratory of the Institute for Advanced Computer Studies, University of Maryland, College Park, where she is currently an Associate Research Scientist. Her research has been in the areas of computational and biological visions centered around the interpretation of the scene geometry from multiple views. Her work is published in 30 journal articles and numerous book chapters and conference articles. Her current interest focuses on visual navigation capabilities, which she studies using the tools of robotics, signal processing, and visual psychology.