Reliability Analysis for Global Motion Estimation

Yu-Chia Chung and Zhihai He, *Senior Member, IEEE*

***Abstract*—Global motion estimation (GME) is the enabling step for many important video exploitation tasks. In this work, we focus on indirect GME methods which have low computational complexity. Typically, an indirect GME method has two major steps. The first step is to find point correspondence between frames through local motion search or feature matching. Then, the second step determines global motion parameters using optimal model fitting, such as least mean-squared error (LMSE) fitting or RANSAC. However, due to image noise and inherent ambiguity in point correspondence, local motion estimation often suffers from relatively large errors, which degrade the performance and reliability of GME. In this work, we propose a method to characterize the reliability of local motion estimation results and use this reliability measure as a weighting factor to determine the importance level of each local motion estimation result during global motion estimation. Our simulation results demonstrate that the proposed scheme is able to significantly improve the accuracy and robustness of global motion estimation with a very small computational overhead.**

***Index Terms*—Motion estimation, reliability analysis, RANSAC, video registration.**

I. Introduction

GLOBAL motion estimation (GME) is the enabling step for many important motion imagery data exploitation tasks, including video registration, moving object detection, tracking, geo-location, and scene understanding [1], [2]. Global camera motion estimation and compensation has also been used in video coding to stabilize images and improve coding efficiency [3], [7].

In a video sequence which experiences global camera motion, two video frames are related by a perspective transform. More specifically, let (x, y) be the pixel position of a point object in frame I_n, where n is the frame index. With global camera motion, this point object moves to a new pixel location in frame I_{n+1}, denoted by (X, Y). The relationship between (x, y) and (X, Y) is given by the following global motion equations:

$$\begin{bmatrix} X \cdot W \\ Y \cdot W \\ W \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

Manuscript received March 24, 2009; revised June 26, 2009. First published July 21, 2009; current version published August 26, 2009. This work was supported in part by the National Institute of Health under Grant 5R21AG026412. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lap-Pui Chau.

The authors are with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 (e-mail: YorkChung@gmail.com; HeZhi@missouri.edu).

Color versions of one or more figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2009.2028101

or, equivalently

$$X = \frac{ax + by + c}{gx + hy + 1}$$
$$Y = \frac{dx + ey + f}{gx + hy + 1}. \tag{2}$$

The objective of global camera motion estimation is to determine the camera model parameters $\{a, b, c, d, e, f, g, h\}$.

There are two basic approaches to global motion estimation (GME), *direct* and *indirect* GME. Direct GME methods determine the global motion parameters by minimizing the prediction error between corresponding pair of pixels in two frames using gradient search or other iterative methods [4]–[6]. For example, in [4], Dufaux and Konrad compute a coarse estimate of the translation component by an -step matching, and then iterate a gradient descent to obtain and refine the eight parameters of a perspective model. Realizing that global motion estimation based on gradient search is too computationally intensive [6], Keller and Averbuch [5] propose to perform the estimation only on a small selective subset of image pixels (called *dominant pixels*) using an interpolation-free computation of global motion. In [3], Su *et al.* attempt to minimize the fitting error between input motion vectors and motion vectors generated by the estimated motion model using a Newton-Raphson method. Indirect GME methods typically consist of two stages. In the first stage, local motion estimation is performed to find point or feature correspondence between two neighboring frames [2], [8]. The second stage determines the global camera motion parameters based on this local point correspondence using model-based fitting [7] or consensus methods (e.g., RANSAC) [9], [10].

Direct GME methods often have high computational complexity. This is because iterative optimization and frequent image warping needed in direction GME are computationally intensive operations [6]. Indirect GME approaches have relatively low computational complexity and implementation cost. It has been extensively used in various global camera motion estimation, image stabilization, and registration schemes [2]. However, indirect GME approaches often suffer from relatively large estimation errors (or robustness) because the local motion estimation in their first stage is often unreliable due to the inherent ambiguity in local motion matching [10].

To address this large error and unreliability issue in local motion matching, indirect GME approaches often resort to optimal model fitting or consensus methods, such as LMSE (least mean squared error) or RANSAC. Here, they assume that a majority of local matching results are close to their true values and they differ by some additive noise. In this case, noise is suppressed during model fitting and outliers are removed by consensus. However, this assumption may

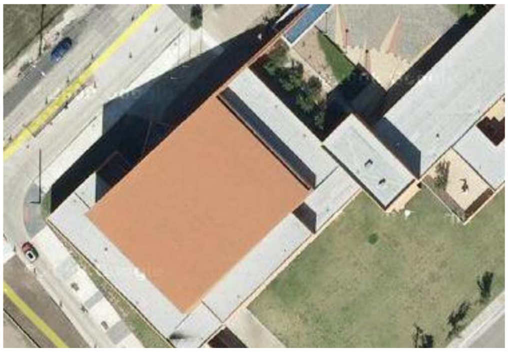


Fig. 1. Aerial image of a building.

not hold in practice. For example, when the video scene is lack of distinctive image features, a majority of local matching results are noisy and unreliable. In this case, the model fitting and consensus approaches may lead to incorrect estimation of global motion parameters. To see this, let us consider the example shown in Fig. 1, an aerial video of a building. If we perform block-based local motion estimation on this video, blocks in these flat image regions and on straight edges will be matched to many blocks in its neighborhood. Therefore, their local motion matching results will be noisy and unreliable. Only those blocks with distinctive features (such as corners and texture patterns) can have reliable local matching. However, the fraction of this type of blocks with reliable local motion estimation is very small. If we apply conventional model fitting and consensus methods, such as LMSE or RANSAC, on these local motion matching results, we will obtain incorrect global motion parameters.

Besides this accuracy and reliability issue, existing methods based on model fitting and consensus treat all local motion matching results to be equally important and do not consider the specific characteristics of each local motion matching. As we know, images are nonstationary data and image characteristics may change dramatically from one region to another. Local motion matching within different image regions may have different level of accuracy and reliability. Therefore, there is a need to develop a model to characterize the local motion matching process and measure its reliability so as to modulate their importance levels during the overall global motion estimation.

To address the above issues, we propose a method to characterize the reliability of local motion estimation results. This reliability measure is then used as a weighting factor to determine the importance level of each local motion estimation result during global camera motion estimation. Our simulation results demonstrate that the proposed reliability analysis and weighting scheme significantly improves the GME performance at a very small computational cost.

The rest of the paper is organized as follows. The reliability analysis of local motion matching is presented in Section II. In Section III, we apply the reliability analysis results to improve the accuracy and robustness of global motion estimation. Experimental results are presented in Section IV. Section V concludes the paper.

Fig. 2. Illustration of reliability.

II. Reliability Analysis of Local Motion Matching

In this section, we develop a method to characterize the reliability of local motion estimation. Let us consider block-based local motion matching and let $\{B^m | 1 \leq m \leq M\}$ be a set of image blocks. For each block, during local motion estimation, we find L best matches in the previous frame, denoted by

$$\Lambda = \left\{ \left(V_j^m, d_j^m\right) | 1 \leq j \leq L \right\}$$

where $V_j^m = (\dot{x}_j^m, \dot{y}_j^m)$ represents the corresponding motion vector and the d_j^m corresponding matching distance. In this work, we use SAD (sum of absolute difference) as our distance metric. Let V^m be the motion vector determined by the local motion estimation which has the minimum matching distance. Note that V^m might not be the true motion vector due to image noise and inherent ambiguity in local motion matching. We define

$$d_-^m = \min_j d_j^m, \quad d_+^m = \text{mean}\left(d_j^m\right).$$

Let

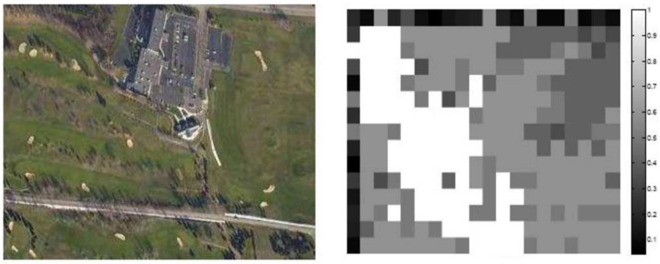
$$d_0^m = d_-^m + \alpha \cdot \left(d_+^m - d_-^m\right)$$

where α is a threshold value between 0 and 1. By default, we set $\alpha = 0.1$. Here, α can be considered as the level of image noise and degree of ambiguity. Fig. 2 shows two cases of local motion estimation within a neighborhood where a minimum distance is found. In case (A), the distance at the minimum location is distinctively smaller than those in its neighborhood. While in case (B), the minimum distance is not distinctively smaller or there might be multiple minimum locations. This implies that this minimum location or the estimated motion vector is not reliable because there are many other locations or motion vectors with similar matching performance. For the image in Fig. 1, those blocks with corner or texture patterns belong to case (A) while those blocks in flat image regions belong to case (B).

We now define

$$\Lambda_- = \left\{ \left(V_k^m, d_k^m\right) | d_k^m \leq d_0^m \right\}$$

which is a set of motion vectors whose distance measurements are very close to the minimum d_-^m. If α is properly chosen or matches the image noise level, then the true motion vector will fall in the set $\Lambda_- \cdot \alpha$ also determines the computational cost due to the number of motion candidates. Smaller alpha trades off less computational cost. Nevertheless, large α is redundant to characterize the motion field with drastic electronic noise. Let's



(a) (b)

Fig. 3. Reliability analysis results of local motion estimation.

relabel the elements in by Γ_- index $k, 1 \leq k \leq k_m < L$. We define the reliability measure as

$$\gamma^m = \frac{1}{1 + \eta \sum_{k=1}^{K_m} \|V_k^m - V^m\|_2}. \quad (3)$$

Here, η is a positive penalty rate and $\|\cdot\|$ represents the L_2-norm. We can see that $0 < \gamma \leq 1$. The major motivation for the proposed reliability formulation in (3) is as follows: 1) K_m is total number of local motion matches whose distance is very close to the minimum one. A smaller value of K_m implies a higher probability or less ambiguity for be the true optimum motion vector, even with the existence of image noise and 2) if the variance of the motion vectors in set Λ_- is small, these motion vectors are close to each other, which means that the estimated motion V^m will be close to the true value. Therefore, the overall reliability is also high.

Fig. 3(b) shows the reliability analysis results of local motion estimation on the image in Fig. 3(a). We can see that the blocks in the top-right region are almost flat and their reliability values are very small. Those blocks in the center-left region have very high reliability because they do have distinctive image features. It is interesting to see that blocks in the parking lot (top-center) area have medium reliability because they look similar to each other and cause some ambiguity during local motion estimation.

III. Global Motion Estimation Based on Reliability Analysis

The reliability measure characterizes how accurate and reliable each block motion estimation is. In this work, we use this reliability as a weighting factor to modulate the importance level of each block during global motion parameter estimation. Let $\{B^m | 1 \leq m \leq M\}$ be the set of image blocks in frame I_n. Let (x_m, y_n) be the center position of block B_m. We apply local motion estimation to each block and find the correspondence of (x_m, y_m) in frame l_{n+1} which is denoted by (X_m, Y_m). According to the global motion equation in (2), we have

$$X_m = \frac{ax_m + by_m + c}{gx_m + hy_m + 1}, \quad Y_m = \frac{dx_m + ey_m + f}{gx_m + hy_m + 1}.$$

It can be rewritten into the following form:

$$P_m \cdot G = Q_m \quad (4)$$

Fig. 4. Example frames of test videos.

where

$$P_m = \begin{bmatrix} x_m & y_m & 1 & 0 & 0 & 0 & -x_m \cdot X_m & -y_m \cdot X_m \\ 0 & 0 & 0 & x_m & y_m & 1 & -x_m \cdot Y_m & -y_m \cdot Y_m \end{bmatrix}$$

$$G = [a, b, c, d, e, f, g, h]^t \quad (6)$$

$$Q_m = [X_m \quad Y_m]^t. \quad (7)$$

In this work, we use a least mean-square error (LMSE) procedure to determine the global motion parameters G which aims to minimize the following square error:

$$E = \sum_{m=1}^{M} [P_m G - Q_m]^t \cdot [P_m G - Q_m]. \quad (8)$$

Note that the reliability of block B^m is γ^m. If we use as a weight, the square error becomes

$$E = \sum_{m=1}^{M} \gamma^m [P_m G - Q_m]^t \cdot [P_m G - Q_m]. \quad (9)$$

Write

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_M \end{bmatrix}, Q = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{bmatrix} \quad (10)$$

Define

$$W = \text{diag}\{\gamma^1, \gamma^1, \gamma^2, \gamma^2, \ldots \gamma^M, \gamma^M\}. \quad (11)$$

The solution to the LMSE problem in (9) is given by

$$G = (P^t W P)^{-1} P^t W Q. \quad (12)$$

It can be seen that the computational overhead of the proposed reliability analysis is very small. The major computation lies in determining the set Λ_-, which has very low computational complexity.

IV. Experimental Results

In this section, we evaluate the performance of the proposed GME algorithm based on reliability analysis. Fig. 4 shows the example frames of our eight test airborne surveillance videos, all at the resolution of 640×480 with 30 frames per second (fps). We use the average Euclidean distance (in pixels) between original and warped pixels [2] to measure accuracy of global motion estimation. We compare our method with the LMSE-based method in [7] and the RANSAC-based global motion estimation method in [10]. Fig. 5 shows the average registration errors (in

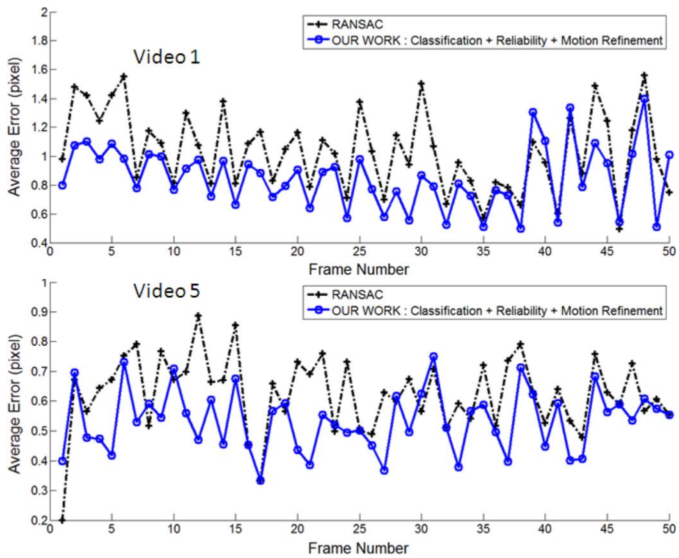


Fig. 5. Average registration errors in pixels of each video frame.

TABLE I
PERFORMANCE COMPARISON WITH RANSAC-BASED GME METHOD

Test Videos	Average Registration Errors (pixels)			Variance of Registration Errors		
	LMSE	RANSAC	This Work	LMSE	RANSAC	This Work
1	3.47	1.04	**0.85**	0.878	0.077	**0.050**
2	0.86	0.56	**0.53**	0.003	0.010	**0.006**
3	1.02	0.56	**0.53**	0.006	0.010	**0.057**
4	1.10	0.55	**0.52**	0.040	0.010	**0.007**
5	1.09	0.63	**0.53**	0.037	0.016	**0.011**
6	1.42	0.63	**0.54**	0.051	0.039	**0.008**
7	0.99	0.51	**0.50**	0.034	0.010	**0.009**
8	0.70	0.44	**0.41**	0.029	0.036	**0.003**

pixels) of each frame from Videos 1 and 5 when the RANSAC method and ours are applied. Table I shows the average registration errors for all eight test videos when these three algorithms are applied. It also shows the variance of registration errors. We can see that, compared to the LMSE and RANSAC-based algorithms, the proposed GME scheme achieve significantly smaller registration errors. The major performance improvement comes from the reliability analysis which effectively characterizes the local motion search process of image blocks and models their importance levels in global motion estimation.

V. CONCLUSION

In this work, we have developed a simple yet efficient method to characterize the reliability of local motion estimation results and use this reliability measure as a weighting factor to determine the importance level of each local motion estimation result during global camera motion estimation. Our simulation results demonstrate that the proposed scheme is able to significantly improve the accuracy and robustness of global motion estimation with a very small computational overhead.

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, pp. 326376–, 1992.
[2] R. P. Wildes, D. J. Hirvonen, S. C. Hsu, R. Kumar, W. B. Lehman, B. Matei, and W. Y. Zhao, "Video georegistration: Algorithm and quantitative evaluation," in *Proc. ICCV 01*, Jul. 2001, vol. II, pp. 343–350.
[3] Y. Su, M.-T. Sun, and V. Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 232–242, Feb. 2005.
[4] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 497–501, Mar. 2000.
[5] Y. Keller and A. Averbuch, "Fast gradient methods based on global motion estimation for video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 4, pp. 300–309, Apr. 2003.
[6] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, pp. 43–77, 1994.
[7] N. Grammalidis, D. Beletsiotis, and M. Strintzis, "Sprite generation and coding in multiview image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 302C311–, Mar. 2000.
[8] A. C. Shastry and R. A. Schowengerdt, "Airborne video registration and traffic-flow parameter estimation," *IEEE Trans. Intell. Transport. Syst.*, vol. 6, no. 4, pp. 391–405, Dec. 2005.
[9] A. M. Buchanan and A. W. Fitzgibbon, "Combining local and global motion models for feature point tracking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
[10] M. Nicolescu and G. Medioni, "A voting-based computational framework for visual motion analysis and interpretation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 739–752, May 2005.