The 31st International Technical Conference on Circuits/Systems,
Computers and Communications (ITC-CSCC 2016)

Histogram-based Non-Iterative Global Motion Estimation

Kyeongbo Kong, Seungjun Shin, and Woo-Jin Song
Department of Electrical Engineering, POSTECH
77 Cheongam-Ro, Nam-Gu, Pohang, Gyeongbuk, Republic of Korea
E-mail : {kkb4723, ymah_89, wjsong}@postech.ac.kr

Abstract: We present a compressed-domain global motion estimation (GME) algorithm that uses a robust estimator. Conventional GME algorithms based on the M-estimator are accurate and powerful, but are slow because they use an iterative process. To speed up the GME, we propose a non-iterative algorithm that exploits an outlier rejection mask and histogram. In simulations the proposed algorithm was about five times faster than the conventional algorithm and had slightly higher estimation accuracy.

Keywords—**Global motion estimation, Motion vector, Robust estimation, Compressed-domain processing, Outlier removal**

1. Introduction

Global motion estimation (GME) in video is used to extract background movement caused by camera motion. GME is useful in applications such as video stabilization, video analysis, object segmentation, and video compression. Global motion can be estimated in either the pixel domain [1] or the compressed domain [2–4]. Pixel-domain GME is usually accurate, but has heavy computation load. Compressed-domain GME is fast because it uses block-based motion vectors (MVs), but has unnecessary MVs (outliers), which should be removed.

Outlier-rejection processes are divided into two groups. The first group is robust estimation, where robust means the ability to reject outliers. These methods first analyze the entire dataset then iteratively remove data that have large error. In this group, M-estimator [2] is widely used. The second group is based on random sampling. These methods first choose a few data randomly and use them to estimate global motion. The process is repeated until a stop condition is satisfied. Helmholtz Tradeoff Estimator (HTE) [4] and Random SAmple Consensus (RANSAC) [5] are members of this group. However, both groups use an iterative approach, which has a high computational load.

In this paper, we present a non-iterative robust GME algorithm that belong to first group of outlier rejection methods. To remove outliers, the proposed method consists of two phases. First phase uses an outlier rejection mask to eliminate outliers of special types such as object, large-difference, and zero or near-zero-magnitude using the outlier rejection mask. The second phase uses M-estimator to assign weights to data remaining after execution of the first phase. For the second phase, we propose a new fitting error that considers both the magnitude of the error and the number of errors. This two-phase process yields accurate results without an iterative process, and can reduce computation load.

2. Transformation model

In [3], four transformation models (translational, geometric, affine, perspective) are described. The perspective transformation model is widely used to estimate global motion. This model has eight degrees of freedom:

$$\mathbf{H} = \begin{pmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{pmatrix}. \tag{1}$$

Using parameter $\mathbf{H}$ of the perspective model, and homogeneous coordinates, a pixel point (x_i, y_i) can be transferred to a new point $(x_i{}', y_i{}')$ as

$$\begin{pmatrix} x_i' \cdot h \\ y_i' \cdot h \\ h \end{pmatrix} = \mathbf{H} \cdot \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}, \tag{2}$$

$$x' = \frac{m_1 x + m_2 y + m_3}{m_7 x + m_8 y + 1}, y' = \frac{m_4 x + m_5 y + m_6}{m_7 x + m_8 y + 1}. \tag{3}$$

If (x_i, y_i) is set as the center of the current frame block that indicates background, and $(x_i{}', y_i{}')$ as a pixel moved using the MV of (x_i, y_i), then $\mathbf{H}$ can be called global motion parameter (GMP). To estimate a perspective model parameter, we first construct a single linear matrix equation as

$$\mathbf{A} \cdot \mathbf{h} = \mathbf{b}, \tag{4}$$

with

$$\mathbf{A} = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x_1' & -y_1 x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y_1' & -y_1 y_1' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & 1 & 0 & 0 & 0 & -x_N x_N' & -y_N x_N' \\ 0 & 0 & 0 & x_N & y_N & 1 & -x_N y_N' & -y_N y_N' \end{pmatrix},$$

$$\mathbf{h} = (m_1, m_2, \ldots, m_7, m_8)^T, \ \mathbf{b} = (x_1{}', y_1{}', \ldots, x_N{}', y_N{}')^T.$$

Then GMP is obtained by the least squares solution of (4) as

$$\mathbf{h} = \left(\mathbf{A}^{\mathbf{T}} \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^{\mathbf{T}} \cdot \mathbf{b}. \tag{5}$$

3. Proposed algorithm

The algorithm (Fig.1) operates frame-by-frame as follows.

3.1 Outlier rejection mask generation

Compressed-domain GME can encounter three kinds of outlier [6]:

- case 1) Object outliers; object motion is usually different from global motion.

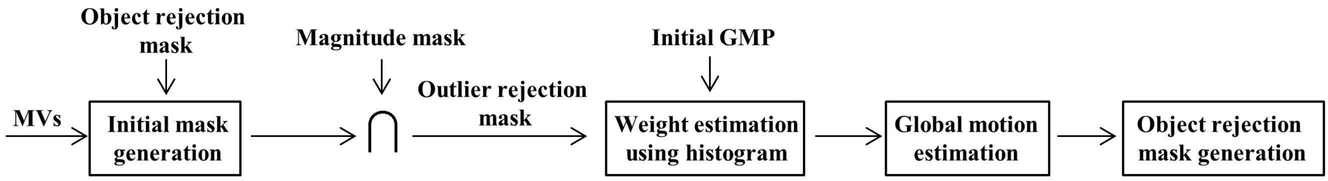


Figure 1. System diagram of histogram-based non-iterative global motion estimation

• case 2) Large-difference outliers; these outliers are caused by matching error such as regions with repetitive or no texture pattern, and boundary regions of a moving object.
• case 3) Zero or near-zero-magnitude outliers; this occurs due to failure of the error minimization step of the block matching algorithm.

Each error type is removed using a specialized procedure. To remove case-1 outliers, we first define an initial mask by using a method similar to that of weight initialization [2]. This mask is obtained by using the object rejection mask of the previous frame. Each block of the current frame is moved to the previous frame by as much as the MV. If moved blocks are overlapped by zero blocks (outliers) of the object rejection mask, then corresponding blocks of current frame are considering to be outliers and their values are set to zero. In the starting frame, we use all one mask because previous frame is not exist.

Most case-2 outliers can be removed by using the proposed magnitude mask, which satisfies

$$(\mu - \sigma) < \|\mathrm{MV}_i\| < (\mu + \sigma), \quad i = 1, 2, \ldots, N, \tag{6}$$

where

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\|\mathrm{MV}_i\|, \quad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\|\mathrm{MV}_i\| - \mu)^2}, \tag{7}$$

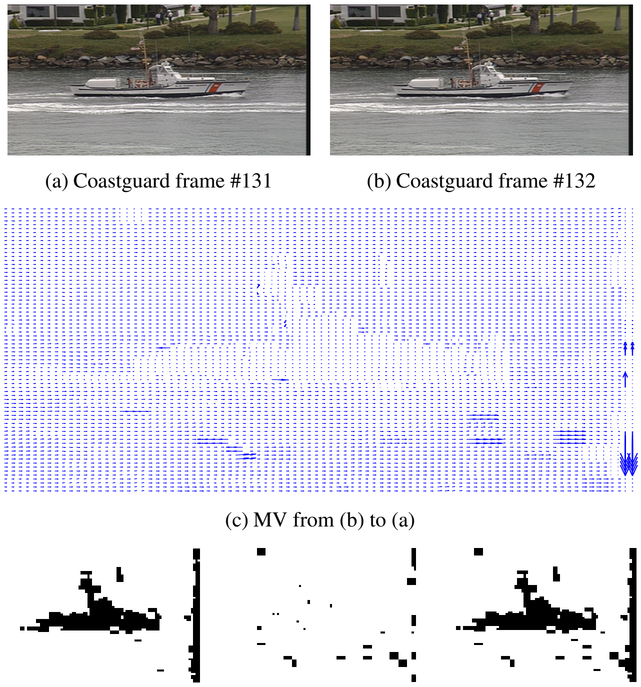
$\|\mathrm{MV}_i\|$ is the magnitude of the i^{th} MV and N is the number of total MVs. This condition means that an MV that is far from the average magnitude is considered to be an outlier.

The outlier rejection mask is defined as

$$\text{initial mask} \cap \text{magnitude mask}. \tag{8}$$

Elimination of case-3 outliers requires information regarding the global motion of current frame. If the magnitude of actual global motion is zero or near-zero in the current frame, then MVs that have magnitude of zero should be accepted as inliers. Therefore, instances in which MV = 0 are considered to be outliers only if MV = 0 for $< 30\%$ of outlier rejection mask blocks that have one value (inlier part).

As an example we use frames from a video of a boat (Fig. 2) to show the process of outlier rejection mask generation which corresponds to frames 131 (Fig. 2a) and 132 (Fig. 2b). MVs from #132 to #131 contain outliers all three types (Fig. 2c). The initial mask removes the boat (object) clearly (Fig. 2d) and the magnitude mask removes large-difference outliers (Fig. 2e). Finally, by generating the intersection of the initial mask and the magnitude mask and eliminating zero-magnitude MVs, we obtain the outlier rejection mask Fig. (Fig. 2f).



(a) Coastguard frame #131 (b) Coastguard frame #132

(c) MV from (b) to (a)

(d) Initial mask (e) Magnitude mask (f) Outlier rejection mask

Figure 2. MV outlier mask from sequence Coastguard.

3.2 Initial GMP estimation

To calculate the weight of each datum, an initial GMP is needed. In a general video sequence, global motion characteristics usually continue over several consecutive frames. Therefore, GMP of the previous frame includes much information about GMP in the current frame, and can be used as its initial estimate. In the starting frame, which has no previous frame, we use the translation model as initial GMP:

$$\begin{aligned} &m_1 = m_4 = 1, \ m_2 = m_5 = m_7 = m_8 = 0 \\ &m_3 = \tfrac{1}{N'}\sum_{i=1}^{N'}\mathrm{MV}x_i, \ m_6 = \tfrac{1}{N'}\sum_{i=1}^{N'}\mathrm{MV}y_i. \end{aligned} \tag{9}$$

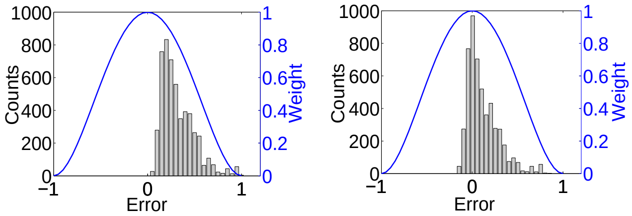
3.3 Weight estimation using histogram

To adjust the influence (weight) of each datum, we use the following function from M-estimator [2]:

$$W(\varepsilon_i) = \begin{cases} \left(1 - \frac{\varepsilon_i^2}{\varsigma^2}\right)^2 & |\varepsilon_i| < \varsigma \\ 0 & |\varepsilon_i| \geq \varsigma \end{cases}, \tag{10}$$

where

$$\varepsilon_i = \left|\mathrm{MV}x_i - \mathrm{MV}'x_i\right| + \left|\mathrm{MV}y_i - \mathrm{MV}'y_i\right|, \tag{11}$$



(a) Conventional error histogram (b) Proposed error histogram

Figure 3. Two error histograms and weight

ς is the tuning constant and ε_i is the fitting error that describes Manhattan distance between MV and MV′, which is the displacement of the center pixel by a GMP. Generally, the initial GMP that represents the global motion of the previous frame does not accurately match the global motion of the current frame; this mismatch causes large initial fitting error. First, a fitting-error histogram is constructed, and error bins are assigned weights. The bin with the largest count usually indicates global motion has error value (0.2) and is therefore assigned a weight < 1 (here, 0.92). This phenomenon does not cause a problem in iterative approaches [1], because the fitting error is adjusted by the iteration step, so global motion is eventually assigned the largest weight, but in non-iterative approaches, this phenomenon degrades the accuracy of outlier detection. To prevent this problem, both the value of errors and the number of errors must be considered.

We propose a new fitting error that is assigned a small value when the Manhattan distance is small and the number of errors is large:

$$\varepsilon_i{}'' = \varepsilon_i{}' \times \frac{N' - h\left(\varepsilon_i{}'\right)}{\frac{1}{N'} \sum\limits_{i=1}^{N'} \left(N' - h\left(\varepsilon_i{}'\right)\right)} \text{ with } \varepsilon_i{}' = \left(\varepsilon_i - \tilde{\varepsilon}\right), \tag{12}$$

where $h(\cdot)$ are histogram bin counts, $\tilde{\varepsilon}$ is the mode of fitting error from the histogram, and N' is the number of outlier rejection mask inliers. This new fitting error imposes two major meanings. 1) Subtracting the mode value from the fitting error sets the error of the largest count to zero. 2) Multiplying by the ratio of complement counts to the average of complement counts causes the error to vary inversely with bin counts. (Fig. 3b) shows new fitting error histogram and weight. The largest bin count which indicates global motion has zero error (0.0) and is assigned the largest weight (1.0). Therefore, appropriate weight can be assigned to all blocks, and the speed and peak signal-to-noise ratio are increased without iteration. Because the mode value is subtracted, we suggest the tuning constant $\varsigma = 1$ rather than $\varsigma = \mu_\varepsilon + 1$, where μ_ε is the inlier sample mean in [2].

3.4 Global motion estimation

By using estimated weight, we can construct a single linear matrix equation as

$$(\mathbf{A} \cdot \mathbf{w}) \cdot \mathbf{h} = \mathbf{b}, \tag{13}$$

(a) Magnitude mask (b) Magnitude mask^C (c) Initial mask

(d) Mismatch mask (e) Outlier rejection mask (f) Object rejection mask

Figure 4. System diagram of histogram-based non-iterative global motion estimation

where

$$\mathbf{w} = \begin{pmatrix} w_1 & w_1 & & w_N & w_N \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ w_1 & w_1 & & w_N & w_N \end{pmatrix},$$

w_i is the weight of the i^{th} block. The GMP can be obtained using (5).

3.5 Object rejection mask generation

At the end of the process, an outlier rejection mask, which is used in the next frame, can be generated using the following condition

$$\varepsilon_i{}''' < \varsigma, \tag{14}$$

where $\varsigma = 1$ and $\varepsilon_i{}'''$ is calculated by using estimated GMP and the equation of conventional fitting error (11) because now the estimated GMP is accurate. However, this mask involves not only objects of case-1, but also matching-error outliers of cases 2 and 3 (Section 3.1) which are meaningless in the next frame. Therefore, we use a magnitude mask to eliminate matching-error outliers as

Object rejection mask:

$$\text{outlier rejection mask} \cup \text{mismatch mask}, \tag{15}$$

where

$$\text{mismatch mask} : \text{magnitude mask}^C \cap \text{initial mask}. \tag{16}$$

4. Simulation results

In this section, we present simulation results to evaluate the processing time and Background Peak Signal-to-Noise Ratio (BPSNR) [2] of the proposed non-iterative GME algorithm. Various sequences were used: Coastguard (SIF, 300 frames), Stefan (CIF, 90 frames), Flower Garden (SIF, 50 frames), Table Tennis (SIF, 67 frames), City (CIF, 100 frames), and Foreman (CIF, 300 frames). All sequences were in YUV 4:2:0 format and had frame rate of 30 fps. We mapped all MVs to 4×4 blocks.

We compared NDLT [2], HTE [4], RANSAC [5], and Cascade [7]. First three algorithms are iterative and the last algorithm is non-iterative. Simulations were performed using

Table 1. Mean processing time per frame [ms]

Algorithm	Image						
	Coast-guard	Stefan	Flower Garden	Table Tennis	City	Fore-man	Avg.
Proposed	4.3	4.1	3.4	3.5	4.3	6.4	**4.3**
NDLT	14.4	14.5	17	18.5	20.5	35.2	**20**
HTE	120.9	116	102.1	100.7	129.7	185.5	**125.8**
RANSAC	55.2	62	60.7	52.9	56.4	90.2	**62.9**
Cascade	9	7.7	7.3	7.6	9.8	11.3	**8.9**

Table 2. Background Peak Signal-to-Noise Ratio [1] [dB]

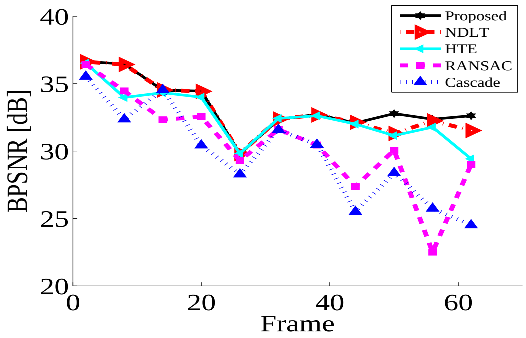
Algorithm	Image						
	Coast-guard	Stefan	Flower Garden	Table Tennis	City	Fore-man	Avg.
Proposed	34.8	33.4	24.4	33	38.7	36.2	**33.4**
NDLT	34.8	33.3	24.4	32.7	38.7	35.9	**33.3**
HTE	34.6	32.9	22.1	32.4	38.5	35.5	**32.7**
RANSAC	34.5	32.2	22.3	30.2	33.5	34.9	**31.2**
Cascade	34.8	31.5	20.3	29.1	35.1	33.1	**30.6**

a desktop computer with 3.50 GHz Intel Core i5-4690 CPU and 8 GB RAM. The proposed algorithm was up to 30 times faster (average $>$ 16 times) than iterative algorithms(NDLT, HTE, RANSAC) and twice as fast as the non-iterative algorithm (Cascade) (Table 1).

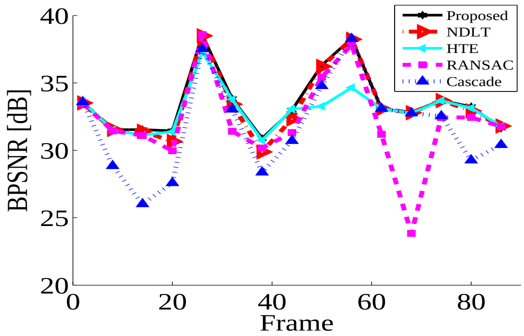
To compare accuracy, we use BPSNR which is used in [2] because global motion is not represented by objects but by background. BPSNR is defined as

$$\mathrm{BPSNR} = 10 \cdot \log_{10}\left(\frac{\sum_{x\in X} 255^2}{\sum_{x\in X} D(x)^2}\right), \qquad (17)$$

where X is set of background pixels and $D(x)$ is intensity difference between previous frame and current frame warped using estimated GMP. Table 2 shows results of various algorithms BPSNR in dB and fig. 5 shows BPSNR values for every 6 frames of Stefan and Table Tennis. The proposed algorithm had slightly higher estimation accuracy than the second-best NDLT algorithm and had an average gain of 2.8 dB over the non-iterative algorithm (Cascade) (Table 2, Fig. 5).



(a) Stefan (CIF, 90 frames)



(b) Table Tennis (SIF, 67 frames)

Figure 5. BPSNR values for every 6 frames

5. Conclusions

we presented a non-iterative robust GME algorithm. To eliminate the need for iteration, we proposed a new outlier rejection mask and new fitting error that exploits an error histogram. In tests on standard test sequences, the proposed algorithm reduce computation load while achieving slightly higher accuracy than existing methods.

6. Acknowledgement

This work was supported by LG Display under LGD-Postech research Program and the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) (IITP-2016-H8601-16-1005) supervised by the IITP(Institute for Information & communications Technology Promotion).

References

[1] A. Krutz, M. Frater, M. Kunter, and T. Sikora, "Windowed image registration for robust mosaicing of scenes with large background occlusions," *Proc. IEEE ICIP06*, Oct. 2006, pp. 353.356.

[2] S. H. Khatoonabadi and I. V. Bajić, "Compressed-domain global motion estimation based on the normalized direct linear transform algorithm," *Proc. ITC-CSCC'13*, 2013.

[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] M. Tok, A. Glantz, M. G. Arvanitidou, A. Krutz and T. Sikora, "Compressed domain global motion estimation using the Helmholtz tradeoff estimator," *Proc. IEEE ICIP.*, pp. 777-780, Sep. 2010.

[5] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381-395, 1981.

[6] Y.-M. Chen and I. V. Bajić, "A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1316.1328, Sep. 2011.

[7] Y.-M. Chen and I. V. Bajić, "Motion vector outlier rejection cascade for global motion estimation," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 197-200, 2010.