A New Decoder Side Video Stabilization using Particle Filter

Masoud Mohammadi, Mahmood Fathi, Mohsen Soryani
Computer Engineering Department
Iran University of Science and Technology
Teheran, Iran
mmohammadi@comp.iust.ac.ir

Abstract— **Machine vision systems which are being extensively used for intelligent transportation applications, such as traffic monitoring and automatic navigation, suffer from image instability caused by environmental unstable conditions. On the other hand, increase in use of home video cameras and need to remove unwanted camera movements which caused by cameraman shaking hands, video stabilization algorithms are being considered. Motion estimation process is the main time consumer phase in video stabilization algorithms. In this paper, by extracting motion vectors from H.264 compressed video, motion estimation process is removed. Moreover, eliminating iterative outlier removal preprocessing and adaptive selection of motion vectors will increase the speed of algorithm. The proposed method was simulated and it was be demonstrated that the extracted motion parameters from motion vectors using particle filter can produce appropriate results for video stabilization problem.**

Keywords- Video stabilization, Particle filter, H.264.

I. INTRODUCTION

The purpose of video processing algorithms which have been developed for image stabilization, is removing unwanted camera motions. The appropriate speed and high accuracy of these algorithms against noisy situations, presence of moving objects and drastic changes in the image depth, are important. On the other hand, different video stabilization subsystems require different levels of image stability. Preserving the generality of the algorithm is also very considerable issue.

Several hardware mechanisms both mechanical and electrical have been developed for the video stabilization. But these mechanical equipments generally, are not accurate and are very massive and heavy. In terms of flexibility, the electrical equipments are extremely limited. Digital solutions can solve these problems well. But main problem of these algorithms is their high time complexity and low efficiency in noisy sequences containing moving objects.

The video stabilization process consists of three essential phases: global motion estimation, intentional motion estimation and motion compensation. The accuracy of parameters obtained in the first stage is very important and also affects other steps' performance strongly. Different motion estimation methods have been proposed.

Feature-based approaches generally show a higher accuracy compared to block matching ones. But using these features or considering any other special conditions for video sequences reduces the generality of the algorithm. The features outgoing from the image area can also make the camera motion estimation process difficult. Reference [4] extracts lane lines and road vanishing point for video stabilization.

Motion vectors, which can be considered as global features, are very useful concept for the motion estimation in various applications. Using MVs instead of any other special feature can increase the algorithm generalization; in addition, it provides the facility for integrating the video stabilization and the video compression subsystems, so the motion estimation process can be removed from one of them. Reference [4,5] and [6] have used MVs which have been obtained directly from the H.264 video sequence, for video stabilization.

Probabilistic approaches that generally become as an estimation problem, contrary to deterministic approaches generally reduce to optimization ones, have high ability to escape from local optima. This fact is the conclusion of randomly search operation. Reference [7,8,9] and [10] have used Kalman filter for motion estimation. But KF is not a powerful algorithm in face of non-linear models with non-gaussian noise. Reference [11,12] and [13] have used Particle filter for motion estimation. All of them have utilized pixel level information, and scale invariant features. Although, SIFT often provides remarkable performance, in this study, we will demonstrate that applying PFs on motion vectors obtained directly from H.264 video sequence, can stabilize unstable video with high accuracy. Fig.1 shows an overview of the proposed algorithm.

The rest of this paper is organized as follows; in section 2, we describe the framework of estimation of inter-frame motion from the H.264 video. Section 3 explains the proposed algorithm for the camera motion estimation using particle filter. Then in Section 4, we discuss the way of dealing with the intentional motion and motion compensation. In section 5, evaluation measurement based on motion vectors field is introduced. Experimental results are presented in Section 6 and Section 7 draws a conclusion.

II. INTER-FRAME MOTION CALCULATION

In the proposed structure, the video stabilizer is integrated with the H.264 coder of decoder side. Prediction vectors which contain motion information between current and reference frames are directly available, in H.264 compression standard. But, main problem is that these prediction vectors do not necessarily relate to consecutive frames. For example, consider the GOP structure in the form of IPB. In this structure, prediction vectors of P-frames depict displacement of this frame relative to the I-frame.

For estimation of the P-frame motion vectors relative to the previous frame, a simple interpolation is applied. So, magnitudes of the prediction vectors are divided to the distance of the current and reference frames. So the MVs which are required for the camera global motion estimation process, will be achieved. While this distance becomes a low value, the estimation would be more accurate. Table-2 shows the way of the interpolation of the motion vectors in different modes of macroblocks.

Many motion models have been proposed in literatures. Despite its simplicity, restricted affine model with four parameters [14], combined with particle filters, has suitable performance in the video stabilization applications. According to this 2D motion model, the displacement of point (x,y) to (x',y') can be represented as follows.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\cos\varphi & -\sin\varphi & T_x \\ \sin\varphi & \alpha\cos\varphi & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (1)$$

Where α, φ, T_x and T_y are scaling, rotation and translation parameters along x and y axis. By extending the equation (1), following relationship exists between the coordinate of i-block center and its corresponding motion vector.

$$\begin{bmatrix} v_x^i \\ v_y^i \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\cos\varphi - 1 & -\sin\varphi & T_x \\ \sin\varphi & \alpha\cos\varphi - 1 & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \qquad (2)$$

By applying equation (2) for each of the image blocks, an equations system will be formed that the number of its equations is more than unknown parameters. For solving this equations system, Least Square Error method can be utilized.

TABLE I. INTERPOLATION OF MVS

Macro Block Type	Previous reference	Next reference	Motion Vector
Intra	---	---	avg(g[n-2], g[n-1])
Forward	n	---	v_F/n
Backward	---	m	$-v_B/m$
Bi-directional	n	m	Avg(v_F/n, $-v_B/m$)

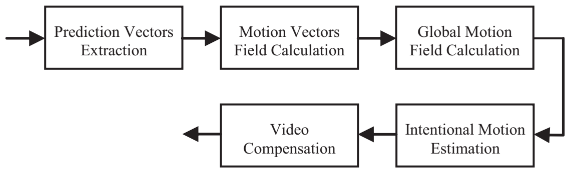

Figure 1. Proposed approach

The LSE method is severely sensitive to outliers in motion vectors. More details can be found in [14].

III. GLOBAL MOTION ESTIMATION USING PARTICLE FILTER

A. Particle Filter

By considering the global camera motion model as a dynamic system, and available MVs field of frame t as a noisy observation of this system, the ME problem can be done by prediction of unknown state of the system in time step t, $\theta_t = [\alpha, \varphi, T_x, T_y]$. State transition and observation models, E_t and O_t respectively, are defined as follows:

$$\theta_t = E_t(\theta_{t-1}, U_{t-1}) \qquad (3)$$

$$Y_t = O_t(\theta_t, N_t) \qquad (4)$$

Where U_t and N_t are the system and the observation noise respectively that have been considered as Gaussian functions. A particle is a weighted sample which can estimate the posterior density [1,2,3]. Each particle, $P_t = \{\hat{\theta}_t, w_t\}$, is determined by its estimated state and corresponding weight. The weight w_t is normalized proportional to the prior probability $P(Y_t|\hat{\theta}_t)$. $Y_t = [v_t^1 \ldots v_t^n]$ is the observations vector at moment t which contains the noisy MVs field information in our desired problem. $Y_{1:t}$ is defined as the set of observations in a time window between 1 and t.

To initialize PFs, K samples, $\{\theta_t^i\}_{i=1}^K$, K samples, $\{U_t^i\}_{i=1}^K$, and K noise samples, $\{N_t^i\}_{i=1}^K$, are drawn from the $P(\theta_{t-1}|Y_{1:t-1})$ and noise probability functions, U_{t-1} and N_t respectively. By substituting these samples in equations (3) and (4), K observations $\{Y_t^i\}_{i=1}^K$ and K estimations for θ_t, can be obtained. Then the weights of particles are calculated as follows:

$$w_t^i \propto \frac{P(Y_t^i|\theta_t)P(\theta_t|\theta_{t-1})}{g(\theta_t|\theta_{t-1}, Y_{1:t})} \qquad (5)$$

Where g(.) is a proposal distribution. In this paper, an approach similar to Sequential Importance Re-sampling filter (SIR) [2,3] is used [15] and an approximation of $P(\theta_t|\theta_{t-1}) \approx g(\theta_t|\theta_{t-1}, Y_{1:t})$ is considered. Therefore, equation (5) can be simplified as,

$$w_t^i \propto P(Y_t^i|\theta_t^i) \qquad (6)$$

Then, the weights of each particle are normalized as,

$$w_t^i = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i} \quad (7)$$

So estimated state can be evaluated as,

$$\hat{\theta}_t = \arg\max_{\theta_t} P(\theta_t|Y_{1:t}) \approx \arg\max_{\theta_t} w_t^i \quad (8)$$

B. Particles population

While the initial state becomes more accurate, the population would be more appropriate. So the smaller noise and the particles number would be required and time complexity of the algorithm would be less. Therefore at first, the motion vectors that their distance from the motion vectors mean is higher than a threshold, (T1=15), are removed. In the next step, this thresholding is done for the motion vectors mode, (T2=20). The selected distance measure is Mahalanobis one that project data correlation and is defined as,

$$D(v_t^i) = (v_t^i - M^t)^T \ \Sigma_t^{-1} \ (v_t^i - M^t) \quad (9)$$

Where, Σ_t is the covariance matrix of the MVs field at the time step t. Subsequent to adaptive selection of MVs, the LS method is used to estimate the initial state. Also the particles are weighted proportional to the sum of Euclidean distance of the estimated MVs field and MVs of observation.

$$w_t^i \propto \frac{1}{\sum_{j=1}^{n} \sqrt{(\hat{v}_{t,x}^{i,j} - v_{t,x}^j)^2 + (\hat{v}_{t,y}^{i,j} - v_{t,y}^j)^2}} \quad (10)$$

Where, $\hat{v}_{t,x}^{i,j}$ and $\hat{v}_{t,y}^{i,j}$ are the components of j-th MV of the estimated MVs field by i-th particle at time t, along x and y axis respectively.

IV. INTENTIONAL CAMERA MOTION ESTIMATION AND MOTION COMPENSATION

For the camera motion smoothing and estimation of the intentional camera motion, a simple and fast causal low-pass filter is used [6]. In fact, this filter is a weighted averaging on the parameters of the camera global motion in a certain period of time T as,

$$\theta_t^d = \sum_{i=1}^{T} w_i * \theta_{t-i}^g \quad (11)$$

Subsequently, by considering a certain frame, for example the first of frame, as the reference frame, motion compensation process can be accomplished as follows:

$$X_{t,i}^s = (A_{t-1}^d * \dots * A_1^d) * (A_{t-1}^g * \dots * A_1^g)^{-1} * X_{t,i} \quad (12)$$

Where $X_{t,i}$ and $X^s_{t,i}$ are the coordinates of i-th pixel in the original and the stabilized frame t. A^g and A^d are the affine matrices of the global and intentional models, respectively.

V. EVALUATION MEASUREMENT OF THE ESTIMATION ACCURACY

With regard to the estimated and ground-truth parameters of the camera global motion and their corresponding MVs field, estimation error can be calculated with mean squared error measurment [14].

$$MSE_t = \frac{1}{n} \sum_{j=1}^{n} ((\hat{v}_{t,x}^j - v_{t,x}^j)^2 + (\hat{v}_{t,y}^j - v_{t,y}^j)^2) \quad (13)$$

The estimated camera global motion will be an accurate model if a low MSE is appeared.

VI. EXPERIMENTAL RESULTS

We evaluate the efficiency of the proposed method through extensive experimental testing. Fig.2 shows a sample result for a video sequence of a road surface. Drastic changes in the depth of video and the uniform texture of the sky and the road's surface make the stabilization process more complex. The needed prediction vectors are extracted by manipulation of JM 14.2 reference software decoder source code. All features are active and the GOP structure is considered as IBP. Fig.3 shows results for a different video sequence. The video sequences stability is clearly observable in both figures.
The motion parameters along x and y axis for the estimated and ground-truth camera global motion are shown in the Fig.4. With regard to the recommended evaluation measurement, the parameters of the ground-truth camera global motion will be required. So, four video sequences with certain motion parameters have been constructed.

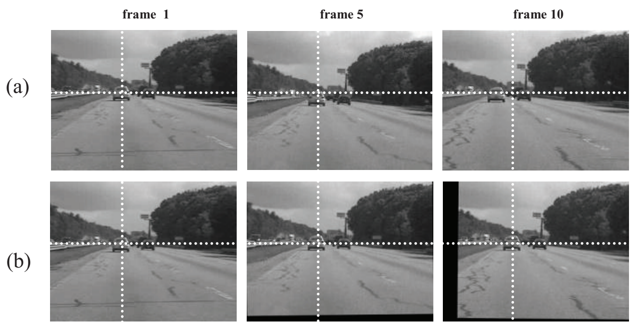

Figure 2. (a) Original and (b) stabilized video sequences

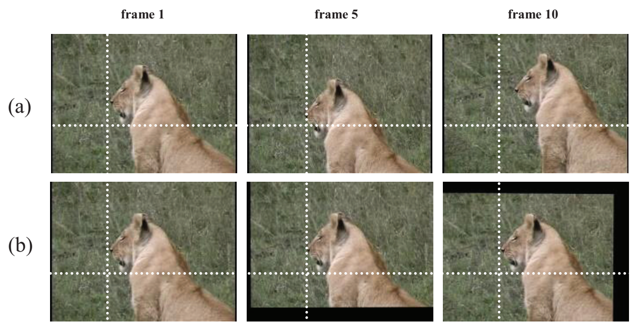

Figure 3. (a) Original and (b) stabilized video sequences

The MSE of the estimated camera global motion for the constructed video sequence is shown in Fig.5.

It is evident that the error of the proposed approach is always less than the LS method. While the video motion becomes more complex, difference of the errors become more distinctive for the LS and proposed methods. Table 2 shows the MSE average of four video sequences.

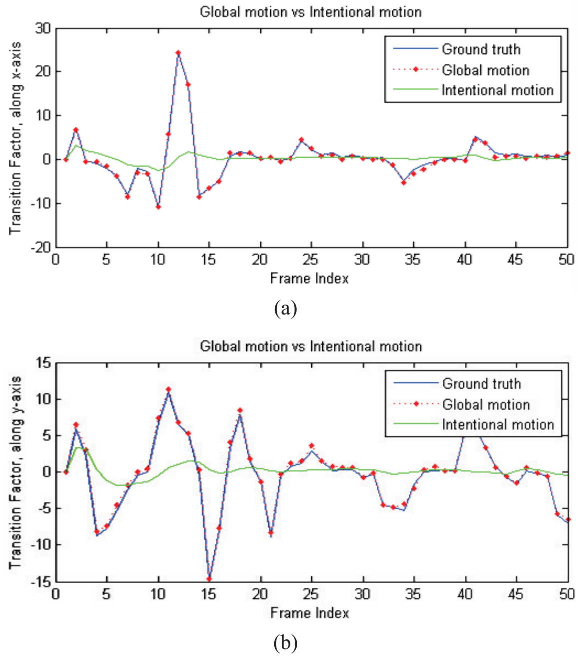

Figure 4: Comparison of the achieved camera motion parameters along two axis, a) x and b) y

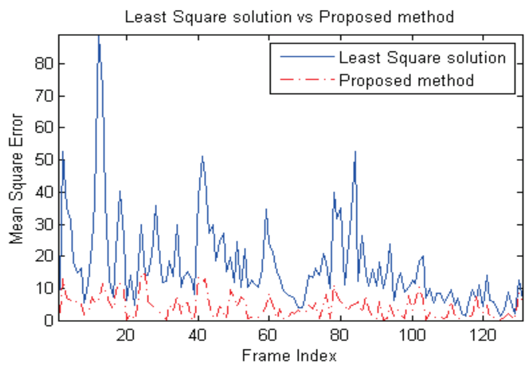

Figure 5: MSE of the proposed and the LS methods

Table 2: Comparison of the MSE average for the LS and the proposed approaches

	Proposed	LS
Sequence 1	4.237	17.197
Sequence 2	3.711	8.419
Sequence 3	6.362	8.681
Sequence 4	0.181	0.195

VII. CONCLUSION

In this paper, a particle filter based approach was presented for the camera global motion estimation which uses extracted prediction motion vectors directly from the compressed video. The motion vectors utilization has increased the generality of the proposed method. By removing of the block or feature matching and any iterative outlier removal process, the proposed approach can be used for online applications. Selection of the motion vectors adaptively, increases the algorithm robustness against the moving objects and drastic changes in the image's depth. Finally, the high performance of the algorithm was demonstrated through various experiments for video stabilization.

REFERENCES

[1] J.H. Kotecha, P.M. Djuric, "Gaussian particle filtering", IEEE Trans. Signal Process., vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[2] S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian Bayesian tracking", IEEE Trans. Signal Processing, vol. 50, no. 2, pp. 174-188, Feb. 2002.

[3] [3] A. Doucet, A.M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen years later", Oxford University Press, Version 1.1, pp. 4-6, Dec. 2008.

[4] Y.M. Liang, H.R. Tyan, S.L. Chang, H.Y.M. Liao, S.W. Chen, "Video stabilization for a camcorder mounted on a moving vehicle", IEEE Trans. on Vehicular Technology, vol. 53, no. 6, pp. 1636-1648, Nov. 2004.

[5] Y.C. Peng, H.A. Chang, H.H. Chen, C.J. Kao, "Digital image stabilization and its integration with video encoder", IEEE Consumer Communications and Networking Conference, pp. 544-549, Jan. 2005.

[6] H.H. Chen, C.K. Liang, Y.C. Peng, H.A. Chang, "Integration of digital stabilizer with video codec for digital video cameras", IEEE Trans. on Circuits and Systems for Video Technology, vol. 17, no. 7, pp. 801-813, July 2007.

[7] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation", IEEE Trans. on Consumer Electronics, Nov. 2003.

[8] Z. Pan, C. Ngo, "Selective object stabilization for home video consumers", IEEE Trans. on Consumer Electronics, vol. 51, no. 4, Nov. 2005.

[9] A. A. Yeni, S. Erturk, "Sast digital image stabilization using one bit transform based sub-image motion estimation", IEEE Trans. on Consumer Electronics, vol. 51, no. 3, Aug. 2005.

[10] A. Litvin, J. Konrad, W.C. Karl, "Probabilistic video stabilization using kalman filtering and mosaicking", IS&T/SPIE Symposium on Electronic Imaging, pp. 663-674, 2003.

[11] J. Yang, D. Schonfeld, M. Mohamed, "Robust Video Stabilization Based on Particle Filter Tracking of Projected Camera Motion", IEEE Trans. on Circuits and Systems for Video Technology, vol. 19, no. 7, July 2009.

[12] Y. Shen, P. Guturu, T. Damarla, B.P. Buckles, K.R. Namuduri, "Video Stabilization Using Principal Component Analysis and Scale Invariant Feature Transform in Particle Filter Framework", IEEE Trans. on Consumer Electronics, vol. 55, No. 3, Aug. 2009.

[13] J. Yang, D. Schonfeld, C. Chen, M. Mohamed, "Online Video Stabilization Based On Particle Filters", IEEE International Conference on Image Processing, pp. 1545-1548, Oct. 2006.

[14] S. Nikitidis, S. Zafeiriou, I. Pitas, "Camera Motion Estimation Using a Novel Online Vector Field Model in Particle Filters", IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 8, Aug. 2008.

[15] S.K. Zhou, R. Chellappa, B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters", IEEE Trans. Image Process., vol. 13, no. 11, pp. 1491-1506, Nov. 2004.