A probabilistic approach for video stabilisation in compressed domain

M Mohammadi*, M Soryani and **M Fathy**
Iran University of Science and Technology, Tehran, Iran

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 environment unstable conditions. On the other hand, by increasing the use of home video cameras which sometimes need to remove unwanted camera movement, which is created by cameraman shaking hands, video stabilisation algorithms are being considered. The video stabilisation process consists of three essential phases: global motion estimation, intentional motion estimation and motion compensation. Motion estimation process is the main time consuming part of global motion estimation phase. Using motion vectors extracted directly from MPEG compressed video, instead of any other special feature, can increase the algorithm generality. In addition, it provides the facility for integrating video stabilisation and video compression subsystems and removing the block matching phase from video stabilisation procedure. Elimination of any iterative outlier removal preprocessing and adaptive selection of motion vectors has increased speed of the algorithm. Although deterministic approaches are faster than the related probabilistic methods, they have essential problems in escaping from local optima. For this purpose, particle filters, the ability of which is considerable when submitted to non-linear systems with non-Gaussian noises, are utilised. Setting the parameters of the particle filter using a fuzzy control system reduces the incorrect intentional camera motion removal. The proposed method is simulated and applied to video stabilisation problem and its high performance on various video sequences is demonstrated.

Keywords: video stabilisation, camera motion estimation, fuzzy control system, particle filter, motion vectors, MPEG

1 INTRODUCTION

The purpose of video processing algorithms which have been developed for image stabilisation, 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 stabilisation subsystems require different levels of image stability. Preserving the generality of the algorithm is also a very considerable issue.

Several hardware mechanisms, both mechanical[1,2] and electrical,[3] have been developed for video stabilisation. But the mechanical equipment generally is not accurate and is very massive and heavy. In terms of flexibility, the electrical equipment is extremely limited. Digital solutions[3–6] can solve these problems well. But the main problem of these algorithms is their high time complexity and low efficiency in noisy sequences containing moving objects.

The video stabilisation process consists of three essential phases: global motion estimation, intentional motion estimation and motion compensation.[6] In the first step, universal movement of the camera

The MS was accepted for publication on 9 June 2012.
* *Corresponding author: Masoud Mohammadi, Iran University of Science and Technology, Tehran, Iran; email: mmohammadi@comp.iust.ac.ir*

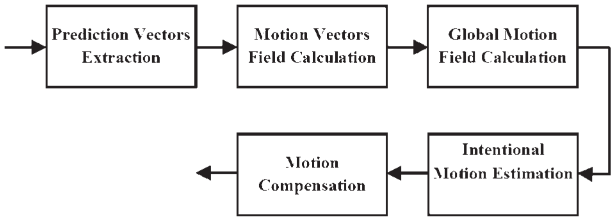
that includes the jerky motion is extracted. Then in the next step, the movements caused by the camera vibration will be separated from desired camera motion. The last step eliminates the unwanted camera motion and generates the stable video sequence.

The accuracy of parameters obtained in the first stage is very important and also strongly affects the performance of other steps. Different motion estimation methods[7,8] have been proposed. Feature-based approaches[9,10] generally show a higher accuracy compared to block matching ones.[11,12] 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 7 extracts lane lines and the road vanishing point for video stabilisation. Reference 10 uses simple features, edges and corners, to viewfinder alignment.

Motion vectors (MVs),[13,14] 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 generalisation. In addition, it provides the facility for integrating video stabilisation and video compression subsystems, so the motion estimation process can be removed from one of them. References 11–14 have used MVs, which have been obtained directly from H.264 video sequence,[13,14] for video stabilisation.

Probabilistic approaches that generally become an estimation problem, contrary to deterministic approaches that generally reduce to optimisation ones, have high ability to escape from local optima.[18] This fact is the conclusion of a random search operation. References 15–21 have used Kalman filter (KF)[25,26] for motion estimation, but KF is not a powerful algorithm in face of non-linear models with non-Gaussian noise.[18] References 23–26 have used particle filter (PF)[22–29] for motion estimation. All of them have utilised pixel level information and scale invariant features.[30–32] Although scale invariant feature transform often provides remarkable performance, in this study, we will demonstrate that applying PFs on MVs obtained directly from H.264 video sequence, can stabilise unstable video with high accuracy. Figure 1 shows an overview of the proposed algorithm.

After extraction of global camera motion model, a way to estimate the desired camera motion should be adopted. For this purpose, different methods are used: Ref. 36 uses a curve fitting method, Ref. 5 implements a fuzzy logic and Refs. 11–13 perform a low-pass filter (LPF) to extract the model of intentional camera motion. These methods which implement the process of video stabilisation by smoothing the parameters of camera global motion, remove portions of desired camera motions wrongly. In such circumstances, cameraman will be felt kind of restrictions during the filming. So it seems that changing the concept of stabilisation from smoothing to camera motion prediction can produce better results. Probabilistic approaches are powerful solutions to predict unknown states of the considered system. References 24 and 38 estimate parameters of camera intentional motion using a KF.

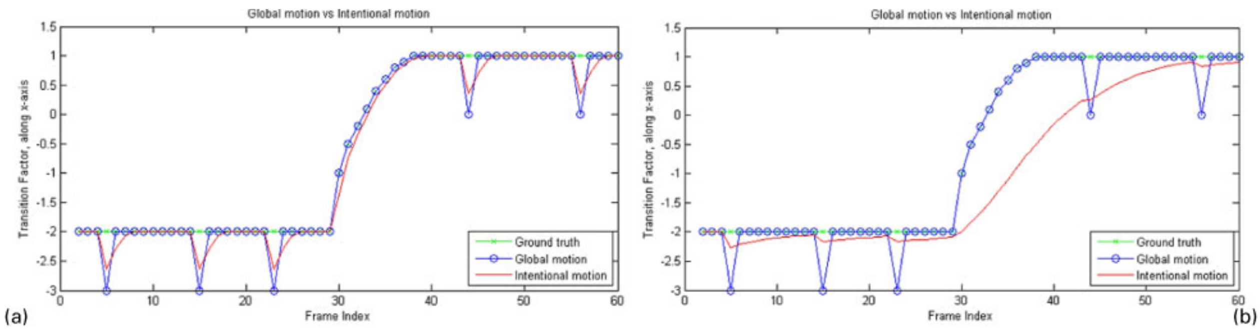


1 Proposed approach overview

Generally, in the paper, we try to provide a video stabilisation algorithm that its final output has best matching with desired camera motion. Using a fuzzy control system instead of a complicated crispy mathematical formulation simplifies the parameters of PF; in addition, the smoothing ability of the PF can be increased. In summary, the novel contributions of this paper are the following:

- extraction of the required MV fields from prediction vector fields directly available in compressed video sequence
- determination of camera global motion using PF that is entirely based on information available in MV level
- elimination of any iterative MV outlier removal by a fast thresholding process that has a good performance on videos that have drastic changes in depth of scene
- using the PF and its online smoothing ability to estimate the intentional camera motion
- specifying the parameters of the PF by performing a simple fuzzy control system
- offering a fully probabilistic approach based on PF for video stabilisation.

The rest of this paper is organised as follows: in Section 2, the basis of video stabilisation strategies



2 The estimation of T_x in the intentional model, using the LPF with (a) $T=3$ and (b) $T=30$

and their disadvantages are presented. In Section 3, we describe the framework of estimation of inter-frame motion from H.264 videos. Section 4 explains the proposed algorithm for estimation of the camera global motion using PF. In Section 5, we discuss the way of dealing with the intentional motion. Section 6 represents a short description for the compensation algorithm. In Section 7, evaluation criteria for the estimation accuracy based on MV field are introduced. Experimental results are presented in Section 8 and Section 9 draws a conclusion.

2 PROBLEM STATEMENT

Many motion models have been proposed in literatures. A 2D translation model with two parameters,[39] a 2D rigid model with four parameters,[40] a 2D affine model with six parameters,[41] a 2.5D model with seven parameters[42] and a 3D model with nine parameters,[43] are instances of these models. Despite its simplicity and speed, restricted affine model with four parameters,[18] combined with PFs, has suitable performance in many video stabilisation applications. According to this 2D motion model, the displacement of point (x,y) to (x',y') can be represented as follows:

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

where α, φ, T_x and T_y are scaling, rotation and translation parameters along x and y axes. By extending equation (1), the following relationship exists between the coordinate of i-block centre and its corresponding MV:

$$\begin{bmatrix} v_x^i \\ v_y^i \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\cos\varphi - 1 & -\alpha\sin\phi & T_x \\ \alpha\sin\phi & \alpha\cos\varphi - 1 & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2}$$

By applying equation (2) for each of the image blocks, an equation system will be formed that the number of its equations is more than unknown parameters. For solving this equation system, least square error (LSE) method[18] can be utilised. The LSE method is severely sensitive to outliers in MVs. More details can be found in Ref. 18. It will be shown that the PF can obtain more accurate parameters for the camera global motion.

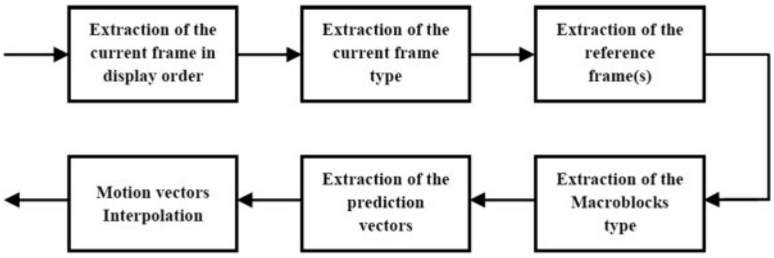
Using the camera global motion model and according to selected motion model, unwanted camera motion must be separated from desired camera motion. Since the frequency of the unwanted camera movement is usually higher than the intentional camera motion,[15,17] a LPF can be utilised for removing the unwanted camera movement from the camera global motion. The proposed approach can be used also for offline applications, but because the proposed algorithm is applied for online applications, only information of the current frame and the former frames will be available and is compared with a causal LPF. This filter is, in fact, a weighted averaging operation on the global motion in a certain period of time T as:

$$\theta_t^d = \sum_{i=1}^{T} w_i \theta_{t-i}^g \tag{3}$$

However, the problem of this method is that if the window size T is considered as a small value, according to Fig. 2a, the power of smoothing in this filter is severely reduced. Subsequently, a large part of vibrations will not be removed. In contrast, with the greater window size (Fig. 2b), intentional motion estimation will not be done appropriately in sudden changes in the parameters of the camera movements. In the rest of the paper, the proposed methods for estimation of the camera global and intentional motion that are based on PF will be described.

3 INTER-FRAME MOTION CALCULATION FROM H.264

In the proposed structure, the video stabiliser is integrated with H.264 coder in the decoder side.



3 Extraction of the motion vectors from compressed video sequence

Prediction vectors which contain motion information between current and reference frames are directly available, in H.264 compression standard. But the main problem is that these prediction vectors do not necessarily relate to consecutive frames. For example, consider the Group of Pictures, 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 MVs 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. The proposed algorithm for extraction of the MVs from compressed video sequences can be seen in Fig. 3.

After extraction of the required information, according to the interpolation methods in Table 1, similar to proposed methods in Ref. 15, MV field can be produced. Using this method, inter-frame MV field can be calculated accurately, but another problem still stays. Block matching process that is done in the H.264 encoder, only considers similarity and does not consider any motion information. Thus, the earned MVs do not necessarily show the macro-block movements. These MVs will be found as outliers in the MV field that are usually originated from uniform textures and changes in scene lighting. Moving objects can also be considered as other sources of outliers.

In the PF algorithm, which is used for camera global motion estimation, while the initial state becomes more accurate, the population would be more appropriate. So the smaller noise and the particle number would be required and time complexity of the algorithm would be less. Therefore, at first, the MVs that their distances from the MV mean are higher than a threshold, T1, are removed. In the next step, this thresholding is performed for the MV mode, T2. The selected distance measure is Mahalanobis one that projects data correlation and is a scale invariant distance measurement. Mahalanobis distance measurement can be evaluated as:

$$D\left(v_{\mathrm{t}}^{\mathrm{i}}\right)=\left(v_{\mathrm{t}}^{\mathrm{i}}-M^{\mathrm{t}}\right)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{t}}^{-1}\left(v_{\mathrm{t}}^{\mathrm{i}}-M^{\mathrm{t}}\right) \tag{4}$$

where $\boldsymbol{\Sigma}_{\mathrm{t}}$ is the covariance matrix of the MV field at the time step t. Subsequent to adaptive selection of MVs, the least square (LS) method is used to estimate the initial state.

4 GLOBAL MOTION ESTIMATION USING PF

By considering the global camera motion model as a dynamic system, and available MV field of frame t as a noisy observation of this system, motion estimation can be done by predicting unknown state of the system in time step t, $\theta_{\mathrm{t}}=[\alpha,\varphi,T_{\mathrm{x}},T_{\mathrm{y}}]$. State transition and observation models, E_{t} and O_{t}, respectively, are defined as follows:

$$\theta_{\mathrm{t}}=E_{\mathrm{t}}(\theta_{\mathrm{t}-1},U_{\mathrm{t}-1}) \tag{5}$$

Table 1 Interpolation of MVs

	Distance from		
Macro-block type	Previous reference	Next reference	MV
Intra	—	—	avg($mv[n-2]$, $mv[n-1]$)
Forward	n	—	pv_{F}/n
Backward	—	m	$-pv_{\mathrm{B}}/m$
Bi-directional	n	m	Avg(pv_{F}/n, $-pv_{\mathrm{B}}/m$)

$$Y_t = O_t(\theta_t, N_t) \tag{6}$$

U_t and N_t, which are the system and the observation noise, respectively, have been considered as Gaussian functions. These Gaussian functions are specialised by a constant zero mean and a covariance adjusted to 0.5. The state transition and observation models are also considered as linear functions. A particle is a weighted sample which can estimate the posterior density.[26,32,33] Each particle, $P_t = \{\hat{\theta}_t, w_t\}$, is determined by its estimated state and corresponding weight. The weight w_t is normalised proportional to the prior probability $P(Y_t|\hat{\theta}_t)$. $\mathbf{Y}_t = [v_t^{o,1} \cdots v_t^{o,n}]$ is the observation vector at moment t which contains the noisy MV 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 initialise the PFs, K samples of $\{\theta_t^i\}_{i=1}^K$, K samples of $\{U_t^i\}_{i=1}^K$ 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. It should be noted that in the utilised PF for camera global motion estimation, a fixed number of particles are considered. Obviously, if the particle number is increased, better accuracy will be achieved for the estimated model but the algorithm speed is reduced. According to the experiments, K=50 particles are considered for each iteration of the PF algorithm. It has been shown that with this number of particles, we can be close to the desired accuracy with a plausible speed. The test results of determining the optimal particle number are presented in the following. By substituting these samples in equations (5) and (6), 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:

$$w_t^i \propto \frac{P(Y_t^i|\theta_t)P(\theta_t|\theta_{t-1})}{g(\theta_t|\theta_{t-1}, Y_{1:t})} \tag{7}$$

where $g(\bullet)$ is a proposal distribution. In this paper, an approach similar to sequential importance re-sampling filter[26,32] is used and an approximation of $P(\theta_t|\theta_{t-1}) \approx g(\theta_t|\theta_{t-1}, Y_{1:t})$ is considered.[18] Therefore, equation (7) can be simplified as:

$$w_t^i \propto P(Y_t^i|\theta_t) \tag{8}$$

Therefore, according to available MV field related to the noisy observation MVs and the particle MVs, the weight of estimated model for each particle in frame t is determined as:

$$w_t^i = \prod_{j=1}^{n_t} P(v_t^{i,j}|v_t^{o,j}) = \prod_{j=1}^{n_t} N(v_t^{i,j}; \mu_t^j, \Sigma_t^j) \tag{9}$$

where $P(\bullet)$ is a bivariate Gaussian probability density function, which is defined for each MV. These probability functions are defined as follows:

$$N(v_t^{i,j}; \mu_t^j, \Sigma_t^j) = \frac{1}{2\pi\left(\left|\Sigma_t^j\right|\right)^{1/2}} e^{-1/2\left(v_t^{i,j}-\mu_t^j\right)^T\left(\Sigma_t^j\right)^{-1}\left(v_t^{i,j}-\mu_t^j\right)} \tag{10}$$

n_t shows the number of MVs which will vary in each frame. The covariance matrices and mean vectors are adjusted to the following fix values:

$$\mu_t^j = v_t^{o,j}[v_{t,x}^{o,j}, v_{t,y}^{o,j}] \tag{11}$$

$$\Sigma_t^j = \Sigma = \begin{bmatrix} 0{\cdot}3 & 0 \\ 0 & 0{\cdot}3 \end{bmatrix} \tag{12}$$

where $v_{t,x}^{o,j}$ and $v_{t,y}^{o,j}$ are the horizontal and vertical components of jth MV in the frame t, respectively. Then, the weights of each particle are normalised as:

$$w_t^i = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i} \tag{13}$$

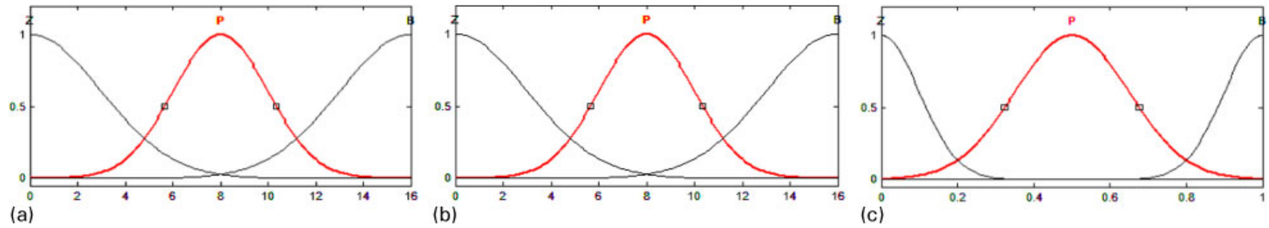
so the estimated state can be evaluated as:

$$\hat{\theta}_t = \arg\max_{\theta_t} P(\theta_t|\,Y_{1:t}) \approx \arg\max_{\theta_t} w_t^i \tag{14}$$

5 INTENTIONAL CAMERA MOTION ESTIMATION BY THE FUZZY PF

In this section, a same PF algorithm, which is employed in the previous stage, is used. Parameters of the camera intentional and global motion in frame t are considered as the dynamic system and its noisy observations, respectively. Unlike before, the parameters of the PF are determined adaptively for each frame. So, a simple fuzzy control system, which is described in the following, is employed.

Roles of the two factors, particle number and variance of the system noise, are very important in generating of the particles population. While the inter-frame variations of the camera global motion parameters become more and continue for a time, a greater number of particles will be required for covering the search space. If not, reducing the number of particles can increase the speed of the proposed algorithm. On the other hand, if the parameters of the global motion have high fluctuations in consecutive frames, it is



4 Definitions of the (a) first input, (b) second input and (c) output of proposed fuzzy control system

necessary to increase the system noise variance to improve the covering of the search space by particles population. Otherwise, the accuracy of the predictions can be increased by decreasing the system noise variance.

Thus by providing a simple fuzzy approach with minimal rules, the particle number and the variance of the system noise are calculated adaptively. Table 2 shows the nine rules that are used in the control system. This control system has two inputs and one output. The first defined input determines the difference of the estimated parameters of the global and intentional camera motion and the second one evaluates the continuation of the fluctuations. The introduced inputs can be defined as:

$$\mathrm{input1}_t = |\hat{\theta}_t - \theta_t| \tag{15}$$

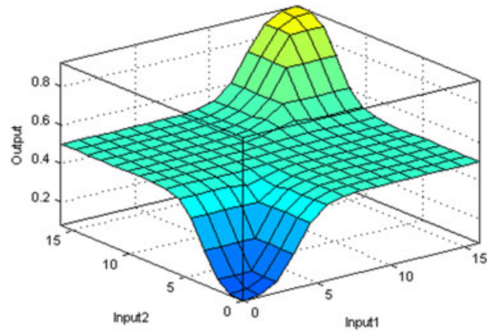
$$\mathrm{input2}_t = |\mathrm{input1}_t - \mathrm{input1}_{t-1}| \tag{16}$$

If the output of this fuzzy control system is shown by F_{out}, the particle number and variance of the system noise can be controlled as:

$$\mathrm{Num}_t \propto F_{\mathrm{out}} \times \mathrm{Num}_{\max} \tag{17}$$

$$\mathrm{Var}_t \propto F_{\mathrm{out}} \times \mathrm{Var}_{\max} \tag{18}$$

The lower and the upper limits of particle number and the noise variance are considered 20–100 samples and 0.1–0.9, respectively. On the other hand, since the camera shaking is generally translation movement, the noise variances of rotation and zoom are considered 10 times smaller than the noise variance of displacement along x and y axes. Specifying the parameters of the PF using the fuzzy control system improves the elimination of the unwanted camera movements and also reduces the incorrect intentional camera motion removal.



5 Changes of the output relative to the inputs of the proposed fuzzy control system

Figure 4 depicts the fuzzy terms of the input and the output in the designed fuzzy control system. According to the mentioned definitions, the changes of the output relative to the inputs can be displayed as the surface in Fig. 5.

After the generation of the particles population, using a method similar to the previous step and considering $P(\theta_t|\theta_{t-1}) \approx g(\theta_t|\theta_{t-1}, Z_{1:t})$, the weight of each particle can be calculated as:

$$w_t^i \propto P(Z_t^i|\theta_t^i) \tag{19}$$

Now, with the assumption of each particle's estimated model $\hat{\theta}_t^i$ as the intentional camera motion and the camera global motion as the system's noisy observation, the weight of each particle can be calculated as follows:

$$w_t^i = \frac{1}{||\hat{\theta}_t^i - Z_t^i||} = \frac{1}{|\hat{\theta}_t^i - Z_t^i|} \tag{20}$$

In other words, for weighting the particles, the Euclidean distance $\|\bullet\|$ is used. Whatever the estimation is closer to the desired state, related particle will have a bigger weight.

Table 2 Rules of the proposed fuzzy control system

		Input 1			
		Z	P	B	
Input 2	Z	Z	P	P	Z=Zero
	P	P	P	P	P=Positive
	B	P	P	B	B=Big-positive

6 MOTION COMPENSATION

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

$$X^{S}_{t,i} = (G^{*}_{t-1} \ldots^{*} G_1)^{*}(D^{*}_{t-1} \ldots^{*} D_1)^{*} X_{t,i} = (G^{*}_{t-1} \ldots^{*} G^{a}_{t-1})^{*}(D^{*}_{t-1} \ldots^{*} D^{a}_{t-1})^{*} X_{t,i} \quad (21)$$

where $X_{t,i}$ and $X^{S}_{t,i}$ are coordinates of ith pixel in the original and the stabilised frame t. $\mathbf{G}_t$ and $\mathbf{D}_t$ are the affine matrices of the global and intentional models, respectively. G^{a}_{t} and D^{a}_{t} are the accumulated global and intentional models for frames 1 to t. In other words, in the first step, the location of the current frame is transformed to the first frame position. Subsequently, by applying the transformation of the desired camera motion, the intentional location of the current frame at moment t can be extracted.

7 EVALUATION CRITERIA FOR THE ESTIMATION ACCURACY

For evaluation of the proposed algorithms, two strategies have been taken. The first method is used when the models of the global and intentional camera motion are completely known. In the same regard, predetermined parameters of the global and intentional camera motion are applied to the frames of the desired video sequence. Using the videos that are made manually, the accuracy of the obtained motion parameters is evaluated using the mean square error (MSE) criterion. If the MV field of the ground-truth model of camera motion is extracted according to equation (2), the MSE criterion in frame t can be defined as follows:

$$\mathrm{MSE}(t) = \frac{1}{N_{\mathrm{mv}}} \sum_{i=1}^{N_{\mathrm{mv}}} \left[(V^{t}_{i,x} - \hat{V}^{t}_{i,x})^2 + (V^{t}_{i,y} - \hat{V}^{t}_{i,y})^2 \right] \quad (22)$$

where V^{t}_{i} and $\hat{V}^{t}_{i}$ are the ith MVs in the corresponding MV field of the ground-truth and the estimated models of the camera motion. The estimated camera motion model will be an accurate model if a low MSE appears. Reference 18 has used the MSE criterion for evaluation of the accuracy of the obtained parameters in their proposed algorithm. But the main problem is that their method cannot be employed in real video sequences.

The second method, which can also be used for real video sequences, is based on this principle that during the stabilisation process, the differences between consecutive frames in the video sequence will be reduced. Accordingly, the inter-frame transformation fidelity (ITF) is used for assessment of the video stability. The ITF criterion can be defined as follows:

$$\mathrm{ITF} = \frac{1}{N_{\mathrm{frame}} - 1} \sum_{t=1}^{N_{\mathrm{frame}}-1} \mathrm{PSNR}(t) \quad (23)$$

$$\mathrm{PSNR}(t) = 10\log_{10} \frac{I_{\max}}{\mathrm{MSE}(t)} \quad (24)$$

$$\mathrm{MSE}(t) = \frac{1}{N_{\mathrm{pixel}}} \sum_{i=1}^{N_{\mathrm{pixel}}} (p^{t}_{i} - p^{t+1}_{i})^2 \quad (25)$$

where N_{frame}, N_{pixel} and $I_{\max}$ are the frame number of the video sequences, number of pixels found in the images and the maximum possible intensity of the pixels. PSNR(t) is the peak signal to noise ratio of frame t and MSE(t) has a same definition as equation (22) that compares the corresponding pixels in two consecutive frames. Higher ITF indicates the more stability of video sequence. References 28 and 44 have used this criterion for evaluation of their proposed algorithm stabilisation ability.

8 EXPERIMENTAL RESULTS

We evaluated the efficiency of the proposed method through extensive experimental testing (the resulting videos can be accessed at http://webpages.iust.ac.ir/masmoh/proj/stabilisation.html). The required prediction vectors are extracted by manipulation of the JM 14.2 reference software decoder source code from H.264 video sequences. All features are active and the Group of Pictures structure is considered as IBP. After extraction of the MV fields, according to the proposed outlier removal procedure, two thresholdings are performed. Although the thresholds have a high tolerance, with regard to numerous tests, these thresholds are set to T1=15 and T2=20. As was said, a PF algorithm with a fixed number of particles is used for estimation of the global camera motion. So the effects of the particle number in accuracy and speed of the proposed algorithm are investigated. The results can be seen in Table 3. The reported execution time does not include the MV extraction phase.

As can be seen, with 50 particles, the desired accuracy can be achieved. More particles do not provide a sensible increment in the accuracy and also decrease the speed of the algorithm. Since the MSE criterion requires the ground-truth model of the camera motion, four video sequences, with certain intentional and vibration models, are manually prepared. So an MATLAB script is provide that randomly [$x \sim N(0,3)$, $y \sim N(0,3)$, $\theta \sim N(0,0.1)$ and $\alpha \sim N(0,0.1)$] select the affine parameters for each frame. Using these values and wrapping a window on a motionless image can construct the desired video sequences. Figure 6 compares the errors of global motions which are estimated by the LS and proposed methods over all frames of video sequence 3.

Table 4 shows the MSE average of the global model estimation for manually constructed video sequences.

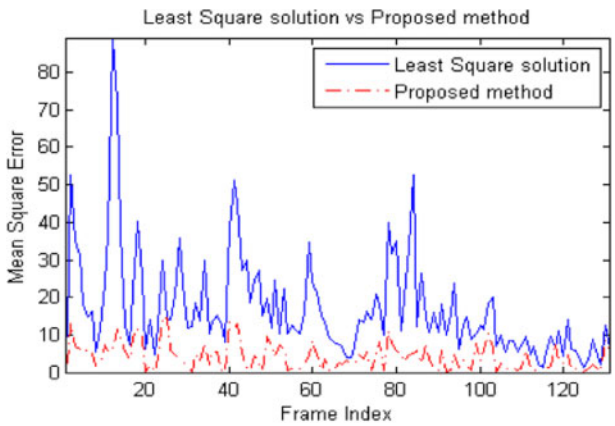
It is evident that the error of the proposed approach is always less than that of the LS method. While the video motion becomes more complex, difference of the errors becomes more distinctive for the LS and proposed methods. Figure 7a shows the estimated camera intentional motion by the proposed algorithm for a video sequence along x axis. As can be observed, camera vibrations have been removed and also the intentional camera motion is tracked properly. Figure 7b illustrates the changes of the system noise variance over all frames. The noise variance is increased in frame 30, so the estimated motion rapidly approaches to the intentional camera motion in next frames.

As another example, the motion parameters estimated in a real video stream along x and y axes for the global and the intentional camera motion are shown in Fig. 8.

Table 5 compares the MSE average of the estimated intentional camera motion in the proposed method and the LPF with window size 20 for constructed video sequences. As can be derived, the proposed method predicts the intentional camera motion with a smaller MSE.

Table 3 MSE average and process time for different numbers of particles

Particle number	MSE	Time (s)
10	8.836	0.2
30	5.722	0.6
50	**4.191**	**0.9**
70	3.924	1.3
90	3.665	2.7



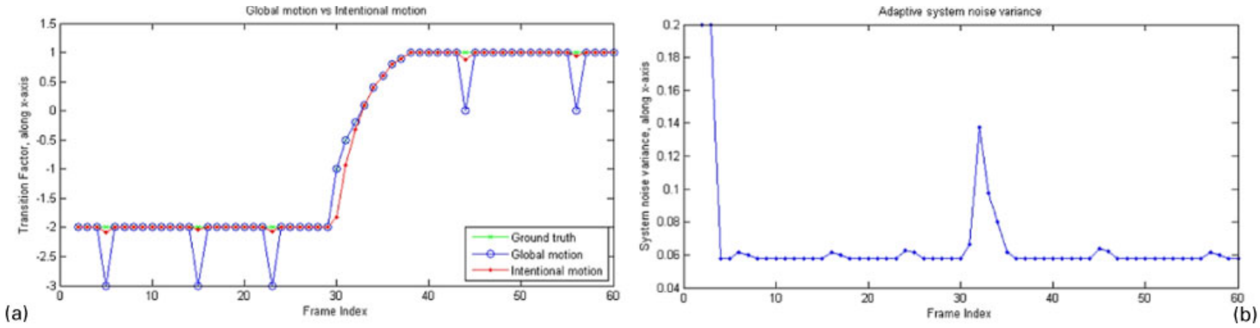
6 MSE of the proposed and the LS methods

Figure 9 shows sample results for several video sequences. Drastic changes in the depth of video, the uniform textures, moving objects and any complicated camera motions, including sudden changes in the intentional camera movement, are handled properly. The video sequence stability is clearly observable in all figures.

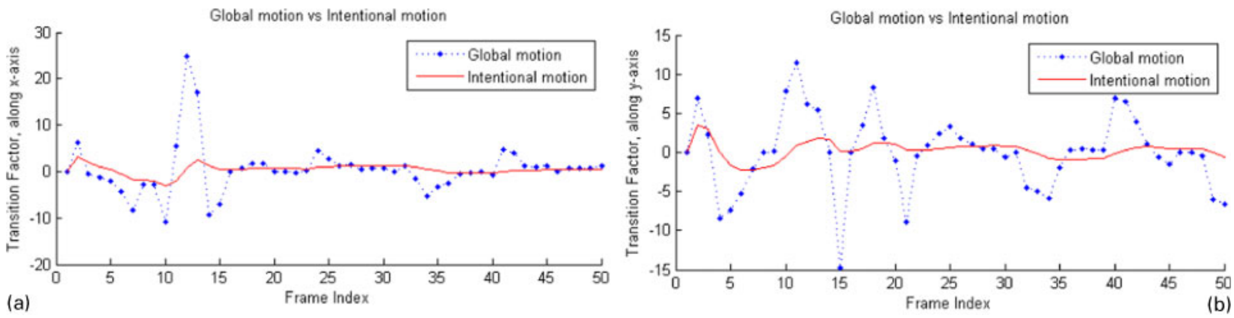
Figure 10 shows the ITF of the 24 video sequences in the original and stabilised mode related to our proposed method and Ref. 15. All resulted video sequences have higher ITF compared to original videos, which demonstrates the final video sequence stability. Reference 15 is a method based on the block matching algorithm integrated with MPEG4 and H.264. The approach proposed in Ref. 15 has shown a high degree of robustness, but its performances are limited by the adopted translational model (e.g. rotation is not taken into account) and the intentional motion estimation algorithm (simple LPF). According to Fig. 10, the proposed approach has higher ITF than[15] for all video sequences. Our algorithm was implemented in a standard MATLAB environment, without using any techniques to speed up the algorithm. Running time was about 1.6 frames per second (0.6, 0.6, 0.2 and 0.2 fps for MV extraction, global motion estimation, intentional motion estimation and motion compensation) on a laptop with a DC 2.5 G CPU and 3 G memory. The image size in all

Table 4 Comparison of the MSE average for the LS and the proposed approaches

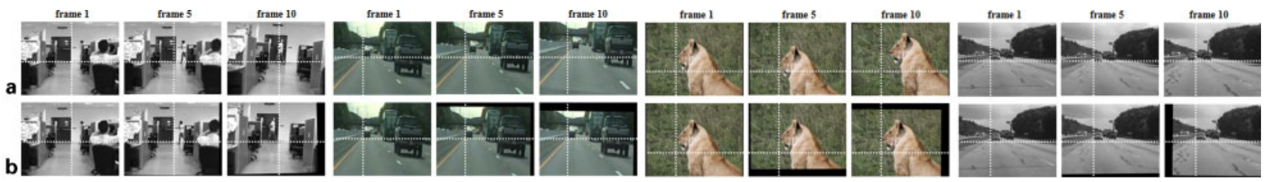
Sequence no.	Proposed method	LS method
1	4.237	7.197
2	3.711	8.419
3	6.362	8.681
4	0.181	0.195



7 (a) Comparison of the achieved camera motion parameters along x axis for one of manually constructed video sequences and (b) adaptive changes of the system noise according to the proposed fuzzy control system



8 Comparison of the achieved models of the global and intentional camera motion for one of video sequences along two axes: (a) x; (b) y



9 Video stabilisation result: (a) original unstable; (b) stabilised video sequences

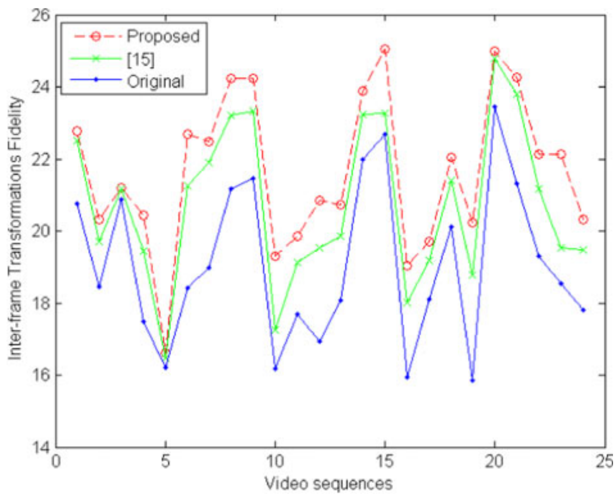
video sequences is 320×240 pixels. Naturally, by hardware implementation, programming in low-level languages and the use of parallel processing concepts, a much higher processing rate can be achieved.

9 CONCLUSION

In this paper, a full PF-based approach was presented for the camera motion estimation which uses extracted

Table 5 Comparison of the MSE average for the LPF and the proposed approaches

Sequence no.	Proposed method	Low-pass filter method
1	0.056	7.717
2	1.933	7.564
3	1.649	5.615
4	0.113	0.147



10 ITF for the original and stabilised video sequences relative to the proposed approach and Ref. 15

prediction MVs directly from the compressed video. The MV utilisation has increased the generality of the proposed method. Removing the block or feature matching and any iterative outlier removal process, speeds up the proposed approach. Selection of the MVs adaptively increases the algorithm robustness against moving objects and drastic changes in the image depth. Using the PF not only increases the accuracy of the estimated camera global motion but also reduces the effects of incorrect intentional camera motion removal. Utilising a fuzzy control system simplifies the determination of the PF parameters. Finally, the high performance of the algorithm was demonstrated through various experiments for video stabilisation.

Possible topics for future work include improving the initial population used in any iteration of the PF algorithms. If the initial population is considered closer to the desired model, the results will be more accurate. Parallel hardware implementation of the proposed approach, integrated with the other intelligent subsystems, is another issue that should be investigated.

On the other hand, although our algorithm has high performance even in critical conditions, it sometimes fails. The presence of homogeneous regions, periodic patterns and sudden illumination changes, if not properly managed,[12] can degrade the performance of proposed method. Previous approaches generally do not work properly in these conditions too.[11] So for the further work, we try to solve these problems.

REFERENCES

1 Vieville, T., Clergue, E. and Facao, P. E. D. S. Computation of egomotion and structure from visual and internal sensors using the vertical cue, Proc. IEEE Int. Conf. on *Computer vision: ICCV '93*, Berlin, Germany, May 1993, IEEE, pp. 591–598.

2 Vermeulen, E. Real-time video stabilization for moving platforms, Proc. 21st Bristol UAV Systems Conf., Bristol, UK, April 2007. Professional Engineering Publishing, pp. 3–4.

3 Egusa, Y., Akahori, H., Morimura, A. and Wakami, N. An application of fuzzy set theory for an electronic video camera image stabilizer. *IEEE Trans. Fuzzy Syst.*, 1995, **3**, 351–356.

4 Chen, C. H., Kuo, Y. L., Chen, T. Y. and Chen, J. R. Real-time video stabilization based on motion compensation, Proc. 4th Int. Conf. on *Innovative computing, information and control: ICICIC '09*, Kaohsiung, Taiwan, December 2009, IEEE CS Press, pp. 1495–1498.

5 Gullu, M. K. and Erturk, S. Fuzzy image sequence stabilisation. *Electron. Lett.*, 2003, **39**, 1170–1172.

6 Zhang, Y., Xie, M. and Tang, D. A central sub-image based global motion estimation method for in-car video stabilization, Proc. 3th Int. Conf. on *Knowledge discovery and data mining: WKDD '10*, Phuket, Thailand, January 2010, IEEE CS Press, pp. 204–207.

7 Dufaux, F. and Moscheni, F. Motion estimation techniques for digital TV: a review and a new contribution. *Proc. IEEE*, 1995, **83**, 858–876.

8 Szeliski, R. Image alignment and stitching: a tutorial. *Found. Trends Comput. Graph. Comput. Vis.*, 2006, **2**, 1–104.

9 Liang, Y. M., Tyan, H. R., Chang, S. L., Liao, H. Y. M. and Chen, S. W. Video stabilization for a camcorder mounted on a moving vehicle. *IEEE Trans. Veh. Technol.*, 2004, **53**, 1636–1648.

10 Adams, A., Gelfand, N. and Pulli, K. Viewfinder alignment. *Comput. Graph. Forum*, 2008, **2**, 597–606.

11 Battiato, S., Bruna, A. R. and Puglisi, G. A robust block based image/video registration approach for mobile imaging devices. *IEEE Trans. Multim.*, 2010, **12**, 622–635.

12 Puglisi, G. and Battiato, S. A robust image alignment algorithm for video stabilization purposes. *IEEE Trans. Circuits Syst. Video Technol.*, 2011, **21**, 1390–1400.

13 Richardson, I. E. G. H.264 and MPEG-4 Video Compression, 2003 (John Wiley & Sons Ltd, Chichester).

14 Ghanbari, M. Standard Codecs: Image Compression to Advanced Video Coding, 2003 (The Institution of Electrical Engineers, London).

15 Chen, H. H., Liang, C. K., Peng, Y. C. and Chang, H. A. Integration of digital stabilizer with video codec for digital video cameras. *IEEE Trans. Circuits Syst. Video Technol.*, 2007, **17**, 801–813.

16 Bhaskar, H. and Bez, H. E. A novel block motion estimation model for video stabilization applications, Proc. 4th Int. Conf. on *Informatics in control, automation and robotics, robotics and automation: ICINCO '07*, Angers, France, May 2007, INSTICC Press, pp. 303–306.

17 Peng, Y. C., Chang, H. A., Chen, H. H. and Kao, C. J. Digital image stabilization and its integration with video encoder, Proc. IEEE Consumer Communications and Networking Conf., Las Vegas, NV, USA, January 2005, IEEE, pp. 544–549.

18 Nikitidis, S., Zafeiriou, S. and Pitas, I. Camera motion estimation using a novel online vector field model in particle filters. *IEEE Trans. Circuits Syst. Video Technol.*, 2008, **18**, 1028–1039.

19 Erturk, S. Digital image stabilization with sub-image phase correlation based global motion estimation. *IEEE Trans. Consum. Electron.*, 2003, **49**, 1320–1325.

20 Pan, Z. and Ngo, C. Selective object stabilization for home video consumers. *IEEE Trans. Consum. Electron*., 2005, **51**, 1074–1084.

21 Yeni, A. A. and Erturk, S. Tast to Fast digital image stabilization using one bit transform based sub-image motion estimation. *IEEE Trans. Consum. Electron*., 2005, **51**, 917–921.

22 Kwon, O., Shin, J. and Paik, J. K. Video stabilization using kalman filter and phase correlation matching. *Lect. Notes Comput. Sci*., 2005, **3656**, 141–148.

23 Erturk, S. Real-time digital image stabilization using Kalman filters. *Real-Time Imag*., 2002, **8**, 317–328.

24 Litvin, A., Konrad, J. and Karl, W. C. Probabilistic video stabilization using kalman filtering and mosaicking, Proc. IS&T/SPIE Symp. on *Electronic imaging*, Santa Clara, CA, USA, January 2003, SPIE, pp. 663–674.

25 Welch, G. and Bishop, G. 'An introduction to the Kalman filter', Techical report 95-041, North Carolina University, Raleigh, NC, USA, July 2006, http://www.cs.unc.edu, 2/26/2011.

26 Arulampalam, S., Maskell, S., Gordon, N. and Clapp, T. A tutorial on particle filters for on-line non-linear/non-gaussian Bayesian tracking. *IEEE Trans. Signal Process*., 2002, **50**, 174–188.

27 Yang, J., Schonfeld, D. and Mohamed, M. Robust video stabilization based on particle filter tracking of projected camera motion. *IEEE Trans. Circuits Syst. Video Technol*., 2009, **19**, 945–954.

28 Shen, Y., Guturu, P., Damarla, T., Buckles, B. P. and Namuduri, K. R. Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework. *IEEE Trans. Consum. Electron*., 2009, **55**, 1714–1721.

29 Sangi, P., Heikkila, J. and Silven, O. Extracting motion components from image sequences using particle filters, Proc. 12th Scandinavian Conf. on *Image analysis: SCIA '01*, Bergen, Norway, June 2001, pp. 508–514.

30 Yang, J., Schonfeld, D., Chen, C. and Mohamed, M. Online video stabilization based on particle filters, Proc. IEEE Int. Conf. on *Image processing: ICIP '06*, Atlanta, GA, USA, October 2006, IEEE, pp. 1545–1548.

31 Karlsson, R. Particle filtering for positioning and tracking applications', PhD thesis, Linkoping University, Sweden, March 2005.

32 Doucet, A. and Johansen, A. M. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later, Version 1.1, pp. 4–6, 2008 (Oxford University Press, Oxford).

33 Kotecha, J. H. and Djuric, P. M. Gaussian particle filtering. *IEEE Trans. Signal Process*., 2003, **51**, 2592–2601.

34 Rong, H., Rongjie, S., Shen, I. F. and Wenbin, C. Video stabilization using scale-invariant features, Proc. 11th Int. Conf. on *Information visualization: IV '07*, Zurich, Switzerland, July 2007, IEEE CS Press, pp. 871–877.

35 Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis*., 2004, **60**, 91–110.

36 Ozbek, F. and Erturk, S. Image sequence stabilization based on adaptive polynomial filtering using the LMS and RLS algorithms. *J. Electr. Electron. Eng*., 2004, **4**, 1233–1242.

37 Leondes, C. T. Fuzzy Theory Systems Techniques & Applications, 1999 (Academic Press, New York).

38 Gullu, M. K., Yaman, E. and Erturk, S. Image sequence stabilisation using fuzzy adaptive Kalman filtering. *Electron. Lett*., 2003, **39**, 429–431.

39 Lee, K. H., Lee, S. H. and Ko, S. J. Digital image stabilizing algorithms based on bit-plane matching. *IEEE Trans. Consum. Electron*., 1998, **44**, 617–622.

40 Morimoto, C. H. and Chellappa, R. Automatic digital image stabilization, Proc. IEEE Int. Conf. on *Pattern recognition: ICPR '96*, Vienna, Austria, August 1996, IEEE. pp. 660–665.

41 Hong, W., Wei, D. and Batur A. U. Video stabilization and rolling shutter distortion reduction, Proc. 17th IEEE Int. Conf. on *Image processing: ICIP '10*, Hong Kong, China, September 2010, IEEE, pp. 3501–3504.

42 Jin, J. S., Zhu, Z. and Xu, G. Digital video sequence stabilization based on 2.5D motion estimation and inertial motion filtering, *Real-Time Imag*., 2001, **7**, 357–365.

43 Zhang, G., Hua, W., Qin, X., Shao, Y. and Bao, H. Video stabilization based on a 3D perspective camera model. *Vis. Comput*., 2009, **25**, 997–1008.

44 Battiato, S., Gallo, G., Puglisi, G. and Scellato, S. SIFT features tracking for video stabilization, Proc. 14th Int. Conf. on *Image analysis and processing: ICIAP '07*, Modena, Italy, September 2007, IEEE CS Press, pp. 825–830.