Digital Video Stabilization System by Adaptive Fuzzy Kalman Filtering

Mohammad Javad Tanakian*
Communication, M.Sc, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
j.tanakian@gmail.com
Mehdi Rezaei
Signal Processing, Assistant Professor, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
mehdi.rezaei@ece.usb.ac.ir
Farahnaz Mohanna
Image Processing, Assistant Professor, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
f_mohanna@ece.usb.ac.ir

Received: 24/Feb/2012 Accepted: 11/Dec/2013

Abstract
Digital video stabilization (DVS) allows acquiring video sequences without disturbing jerkiness, removing unwanted camera movements. A good DVS should remove the unwanted camera movements while maintains the intentional camera movements. In this article, we propose a novel DVS algorithm that compensates the camera jitters applying an adaptive fuzzy filter on the global motion of video frames. The adaptive fuzzy filter is a Kalman filter which is tuned by a fuzzy system adaptively to the camera motion characteristics. The fuzzy system is also tuned during operation according to the amount of camera jitters. The fuzzy system uses two inputs which are quantitative representations of the unwanted and the intentional camera movements. Since motion estimation is a computation intensive operation, the global motion of video frames is estimated based on the block motion vectors which resulted by video encoder during motion estimation operation. Furthermore, the proposed method also utilizes an adaptive criterion for filtering and validation of motion vectors. Experimental results indicate a good performance for the proposed algorithm.

Keywords: Adaptive, Digital Video Stabilization, Fuzzy Filter, Kalman Filter, Motion Estimation, Motion Vector, Video Coding.

1. Introduction

Digital video stabilization (DVS) techniques have been studied for decades to improve visual quality of image sequences captured by compact and light weight digital video cameras. When such cameras are hand held or mounted on unstable platforms, the captured video generally looks shaky because of undesired camera motions. Unwanted video vibrations would lead to degraded view experience and also greatly affect the performances of applications such as video encoding [1-4] and video surveillance [5,6]. With recent advances in

Wireless technology, video stabilization systems are also considered for integration into wireless video communication equipment for the stabilization of acquired sequences before transmission, not only to improve visual quality but also to increase the compression performance [1]. Solutions to the stabilization problem involve either hardware or software to compensate the unwanted camera motion. The hardware-based stabilizers are generally expensive and lack the kind of compactness that is crucial for today's consumer electronic devices [7, 8]. On the contrary, a DVS system that is implemented by software can easily be miniaturized and updated. Consequently, DVS system is suitable for portable digital devices, such as digital camera and mobile phone.

In general, a DVS system consists of two principal units including motion estimation (ME) and motion correction (MC) units. The ME unit estimates a global motion vector (GMV) between every two consecutive frames of the video sequence. Using the GMVs, the MC unit then generates smoothing motion vectors (SMVs) needed to compensate the frame jitters and warp the frames to create a more visual stable image sequence.

According to the motion models being considered, the already proposed global ME techniques for DVS system can roughly be divided into two categories:

(1) Two- dimensional stabilization techniques which deal with translational jitter only [9-20] and

(2) multi-dimensional stabilization techniques which aim at stabilizing more complicated fluctuations in addition to translation [21-25]. Most of the existing algorithms fall into the first category because the translation is the most commonly encountered motion and the complexity of estimating translation parameters is relatively low for real-time stabilization. In the second category, the majority of algorithms [21,22,24] considered a three-dimensional or perspective motion model, while a few algorithms considered affine motion

* Corresponding Author

model [23] or sensors attached to the camera to provide absolute 3D orientation [25].

Regarding to the two-dimensional ME task of DVS systems, most previous approaches attempt to reduce the computational cost by using fast ME algorithms, e.g. gray-coded bit-plane matching [9], two-bit transform [10], multiplication-free one-bit transform [11], Laplacian two-bit transform [12], and binary image matching of color weight [13]. In another approach, the global ME is limited to small, pre-defined regions [16,17]. Such approaches consider DVS and video encoding separately and attempt to trade the accuracy of motion vectors (MVs) for the computational efficiency; nevertheless they improve the computational efficiency at the expense of degradation in the accuracy in ME and thereafter in MC tasks.

Since both the video encoder and the digital stabilizer of a digital video camera use a ME unit, we can integrate digital stabilizer with video encoder [2,4,26] by making the two modules of a digital video camera share a common local motion vectors (LMVs) estimation process, as shown in Fig. 1. ME can take up 90% of the total computation of a digital stabilizer and 50% -70% of a video codec. Combined together, the operation required for ME can consume more than 70% of the total computation [4]. The ME task in video encoders usually is implemented on frame blocks by a block matching process to estimate a MV for each block (BMV).

The ME unit plays an important role in DVS system and its estimation accuracy is a decisive factor for the overall stabilization performance of the system. In video frames with smooth or complex texture regions, the estimated BMVs may not be in coincidence with the real motion of the blocks. Although such LMVs are applicable to the local motion compensation task which is executed in the encoder, they cannot be used for the global motion compensation which is executed by the DVS. These LMVs include some noises that degrade the global ME task. In order to remove the noisy LMVs in these regions some algorithms are proposed in [27-30]. The valid BMVs as LMVs are used for the global ME and MC compensation in next steps.

After global ME, the next essential task of a DVS system is MC in which the unwanted camera jitters are separated and removed from the intentional camera movement. Among the various MC algorithms proposed in the literature, smoothing of the GMV by low-pass filtering is the most popular. For instance, an MV integration method is used in [9,31] which utilizes a first-order infinite impulse response (IIR) low-pass filter to integrate

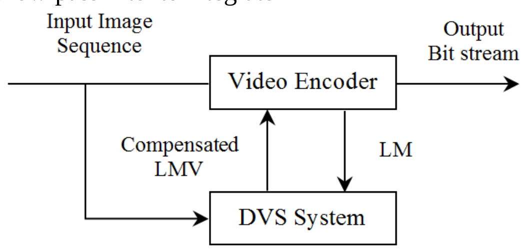


Fig.1 Integration scheme of the video stabilizer and the video encoder.

Differential motion and to smoothen the global movement trajectory. A frame position smoothing (FPS) algorithm, based on smoothing absolute frame positions that achieve successful stabilization performance with retained smooth camera movements, is utilized for MC in [17,32-41]. Off-line discrete Fourier transform (DFT) domain filtering is proposed for FPS-based stabilization in [32].

Kalman filter and fuzzy systems have widely been used in DVS applications [33-41]. A real-time FPS-based stabilizer using Kalman filtering of absolute frame positions has been proposed in [17,33]. In the presented algorithm in [34] two Kalman filters are operated in parallel, one of which is used as a reference filter with a constant high process noise variance and another one is used as stabilization filter in which a fuzzy system lets the process noise variance to be adjusted. In this case, the process noise variance of the stabilization filter is adaptively changed by the fuzzy system according to the residual between the stabilization and reference filter output. Presented DVS in [35] utilizes an fuzzy system in which membership functions (MFs) are optimized to motion dynamics [35]. A membership selective fuzzy stabilization, in which the stabilization system selects between a pre-determined set of MFs according to instantaneous motion characteristics is proposed in [36]. A MF adaptive fuzzy filter based on smoothing of absolute frame position for video stabilization is presented in [37]. In this method initially a short mean filter is applied to raw absolute frame displacement as pre-process, to reduce the dynamic range of the fuzzy system input. Fuzzy stabilization is then achieved through fuzzy correction mapping. In this method output MFs of the fuzzy system are continuously adapted so as to constitute a MF adaptive fuzzy filtering process. It is shown in [38] that the performance of the Kalman filter can be improved by a fuzzy system that improves the Kalman filter output. In presented algorithm in [33,39], the process noise variance in Kalman filter is adjusted adaptively according to the camera motion characteristics.

Almost all fuzzy systems utilized by presented algorithms [34-38] have a similar structure with two inputs as fuzzy inputs and a little difference in fuzzy MFs. In these algorithms, the first input is the difference between the absolute frame displacement and the a priori estimate of the stabilized frame position that achieve by a Kalman filter or another predictor. The second input indicates the change of first input over the last two frames.

Regarding to the MC task of DVS systems, almost all published algorithms try to smoothen the global movement trajectory by a kind of low-pass filtering. An important drawback of the low-pass filtering is that smoothened movement trajectory is delayed with respect to the desired camera displacements. A stricter filtering provides more stabilization at the expense of more trajectory delay and vice versa. More trajectory delay means losing more image content after stabilization.

In this article, we propose a DVS algorithm with new features in ME and MC units. The ME unit estimates a

GMV based on the BMVs which are estimated by the video encoder. Therefore, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. Moreover, in order to improve the accuracy of GMV estimation task an adaptive thresholding algorithm is used to remove the noisy invalid LMVs. The MC unit of the proposed DVS system is an adaptive fuzzy filter that applied on the global motion of video frames to smooth the camera movement trajectory adaptively. The adaptive fuzzy filter is a Kalman filter which is tuned by a fuzzy system adaptively to the characteristics of unwanted and intentional camera motions. The fuzzy system itself is also tuned during operation according to the amount of camera jitters. The fuzzy system uses two inputs which are quantitative representations of the unwanted and the intentional camera movements. Experimental results show a good performance for the proposed DVS algorithm. Also in order to have a generic comparison between our method and some relevant MC algorithms proposed in the literature, Table1 is presented. A good MC unit should remove the unwanted camera motion while tracks the intentional motion without any delay. For this purpose, it should discriminate the unwanted and intentional camera motions while adjust the smoothing filter adaptively according to the amount of unwanted and intentional camera motions. The studied published MC algorithms lack some of these features. For example in the Table1, it is shown that the algorithms presented in [27,39,40] suffer from the lack of discrimination of unwanted and intentional camera motions. Moreover, the proposed adaptive algorithm in [27] suffers from a continuous and well adaptation. They use an adaptive filter with a smoothing factor that is switched between only two values and therefore it leads to undesirable jumps in frame position while our proposed MC algorithm discriminate between the unwanted and intentional camera motions and also uses a fuzzy system to improve the quality of adaptive filter.

The remainder of this article is organized as follows. The details of the proposed video stabilization algorithm are

Table 1: Comparison results for some MC algorithms proposed in the literature and our Method

Criteria	Adaptive Filter	Kalman Filter	Fuzzy System	Discriminating Unwanted and Intentional Motions
Adaptive IIR Filter [27]	Y			
Adaptive Kalman Filter [33]	Y	Y		
Fuzzy Adaptive Kalman Filter [34]	Y	Y	Y	
MF Adaptive Fuzzy Filter [37]	Y		Y	
Recursive Fuzzy System [38]		Y	Y	
Adaptive Kalman Filtering [39]	Y	Y		
Adaptive Kalman Filtering [40]	Y	Y		
Proposed Method	Y	Y	Y	Y

described in Section 2. Some experimental results are presented in Section 3, and the article is concluded in Section 4.

2. The Proposed Method

A flowchart of the proposed DVS system is depicted in Fig.2. The details of the proposed system are described in the sequel.

2.1 Block-Based ME

The block-based ME is used to generate the LMVs. Since the ME is done by the video encoder, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. In this article, to test the proposed DVS system independent of the encoder, a full search ME algorithm with full-pixel resolution is taken for 8×8 blocks over a search range of 33×33 pixel to achieve the BMVs.

The ME algorithm works as follows. First, the current frame is divided into a number of *N×N* blocks and an MV for each block is computed. The resulting MV points to the most correlated reference block in the previous frame within the search area. To measure the goodness of each

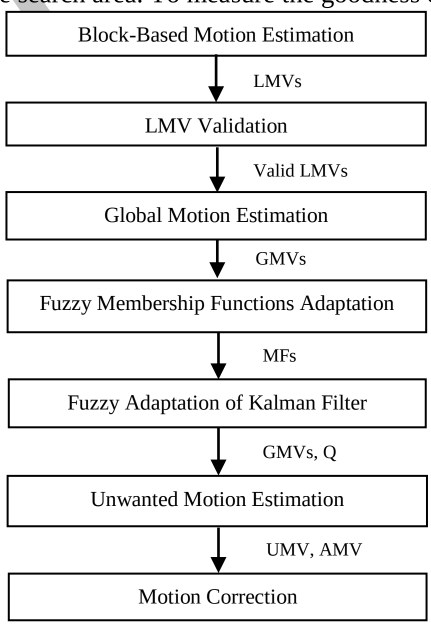


Fig. 2 Flowchart of the proposed DVS system.

candidate MV (x,y), the mean absolute difference (MAD) measure is used as

$$MAD(i,j) = \frac{1}{N^2}\sum_{K=0}^{N-1}\sum_{L=0}^{N-1}|C(x+k, y+k) - R(x+i+k, y+j+l)| \quad (1)$$

where C(x+k, y+L) and R(x+i+k, y+j+L) denote the block pixels in the target frame, and the displaced block pixels in the reference frame, respectively. The candidate MV (i,j) with the smallest MAD is chosen as the MV of the current block according to:

$$V = \arg_{(i,j)} \min MAD\,(i,j),\ -p \le i,j \le p-1 \qquad (2)$$

Where p defines the motion search range.

2.2 LMV Validation

The ME unit plays an important role in DVS system and its estimation accuracy is a decisive factor for the overall performance of stabilization system. Block ME process typically computes some wrong MVs which are not in coincidence to the real motion direction of the blocks. Although, such MVs can be useful for the motion compensation in encoder, they include noise and should not be used for the global motion compensation and video stabilization operations. These LMVs include some noises that degrade the global ME task. The noisy MVs are mostly obtained from two types of regions including: very smooth regions with lack of features and very complex uneven regions [27-30]. Inspiring from the algorithm presented in [27], two qualifying tests, namely “Smoothness Test” and “Complexity Test”, are used to detect and remove the noisy MVs by an adaptive thresholding method as follows. The valid BMVs as LMVs are used for the global ME and MC in next steps.

2.2.1 Smoothness Test

The noisy MVs corresponding to the smooth regions such as sky image are detected by thresholding of the average of MAD as:

$$MAD^{n}_{avg} < th_1, \qquad (3)$$

where MAD^{n}_{avg} denotes the average of calculated MADs within the search area, during ME of n^{th} block. th_1 is also defined as

$$th_1 = MAD^{n}_{min} + T_1 \times Mean(MAD^{n}_{avg}) \qquad (4)$$

Where MAD^{n}_{min} and MAD^{n}_{avg} denote the minimum and the average values of computed MADs, respectively, during ME of n^{th} block within the search area. T_1 is an experimentally defined constant coefficient about 0.45 and $Mean(MAD^{n}_{avg})$ denotes the average of MAD^{n}_{avg}, over all blocks of the frame. In fact the threshold th_1 includes a global average value over the frame plus a margin.

2.2.2 Complexity Test

The noisy MVs corresponding to the complex texture regions are identified by another thresholding as:

$$MAD^{n}_{min} > th_2, \qquad (5)$$

Where threshold th_2 is defined adaptively as:

$$th_2 = T_2 \times Max(MAD^{n}_{min}), \qquad (6)$$

Where T2 is an experimentally defined constant coefficient about 0.45, and $Max(MAD^{n}_{min})$ denotes the maximum value of MAD^{n}_{min}, over all blocks of the frame. According to the equations above, the MAD^{n}_{min} is compared against a portion of its global maximum over a frame.

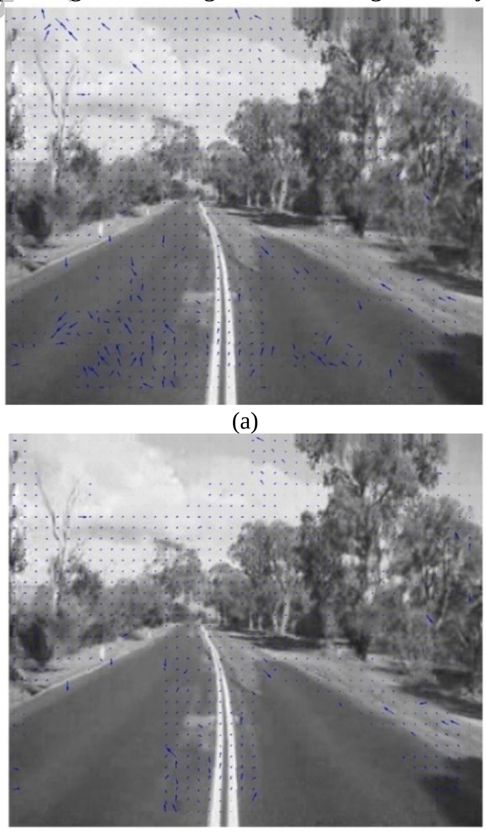
It is notable that MAD is computed during ME by encoder. Therefore, the smoothness test and complexity test have no additional computational complexity cost for the proposed DVS system.

Provided results on different video contents show that, using fixed thresholds for different video contents may cause a remarkable amount of invalid noisy LMVs remain or a notable amount of valid LMVs be removed. To solve this problem, the values of thresholds th_1 and th_2 are adjusted adaptively based on the video content for each frame. Note, if ME is executed by a fast search algorithm rather than full-search algorithm at the encoder, the MADs calculated during ME are used for adaptation of thresholds th_1 and th_2.

Original LMVs and validated LMVs for a sample frame are presented in Fig.3. This figure shows that many noisy LMVs have been removed by the LMV validation process.

2.3 Global ME

The global ME unit produces a unique GMV for each video frame, which represents the camera movement during the time interval of two frames. Since the LMVs obtained from the image background tend to be very similar in both magnitude and direction, we used a clustering process to classify the motion field into clusters corresponding to the background and foreground objects.



(a)

(b)

Fig. 3 Example of noisy LMV removal on a frame of the avenue sequence. (a) Original MVs, (b) Valid MVs.

The global motion induced by camera movement is determined by a clustering process that consists of the following steps.

Step 1) Construct the histogram H of the valid LMVs. The value of H(x,y) is incremented by one each time the LMV(*x,y*) is encountered.

Step 2) As long as the scene is not dominated by moving objects, the cluster corresponding to background blocks has the maximum votes in the clustering process. The position (*x,y*) of the largest cluster or histogram bin is considered as the GMV.

As an example, Fig. 4 shows the largest histogram bin at coordinates (5,12), yields the GMV.

2.4 Unwanted ME and Correction

An estimated GMV may consist of two major components: an intentional motion component (e.g., corresponding to camera panning) and unintentional motion component (e.g., corresponding to handshake). A good MC algorithm should only remove the unwanted motion while maintain and track the intentional motion. Assuming that the unwanted motion is corresponding to the high-frequency components, the proposed algorithm uses a low-pass filter to remove the unwanted motion component. An SMV is resulted by a Kalman filtering on the GMVs that resembles the intentional camera movement. The Kalman filter provides an estimation to the state of a discrete-time process defined as a linear dynamical system as

$$\mathrm{x}(\mathrm{t}+1) = \mathrm{F} \times \mathrm{x}(\mathrm{t}) + \mathrm{w}(\mathrm{t}). \quad (7)$$

Where *F* and *w(t)* represent state transition matrix and the process noise, respectively. The Kalman filter operates

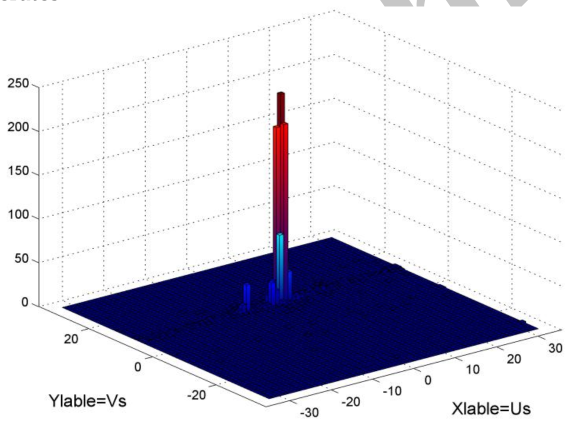


Fig. 4 Clusters of Motion field

using observations defined by the observation system as

$$\mathrm{y}(\mathrm{t}) = \mathrm{H} \times \mathrm{x}(\mathrm{t}) + \mathrm{v}(\mathrm{t}). \quad (8)$$

Where *H* and *v(t)* show measurement matrix and noise, respectively. Process and measurement noise are assumed to be independent of each other, white, and with normal probability distributions: w ~ *N* (0, *Q*) and v ~ *N* (0, *R*).

The Q (process noise variance) in Kalman filter has a direct effect on the operation of stabilizer. A relative small variance value provides expanded stabilization with reduced adaptability to changes in intentional motion dynamics, it means Kalman acts as a strong filter and provides expanded stabilization. While a relative large variance value enables close tracking of intentional camera movements but at the cost of slightly reduced stabilization capabilities, it means Kalman acts as a weak filter and quickly adjusts to changes in intentional motion dynamics. So, we can have an adaptive Kalman filter if the process noise variance of Kalman filter changes automatically during filtering operation. A fixed value of Q not effectively leads to good stabilized image sequences [13]. To avoid the lag of intentional movement and to smooth the unwanted camera motion efficiently, the following fuzzy adaptation mechanism of Q is proposed.

2.4.1 Fuzzy Adaptation of Kalman Filter

The Kalman filter is implemented on the vertical and horizontal components of the GMVs separately. The process noise variance of Kalman filter i.e., Q(n) is adjusted by a fuzzy system continuously for MC of each frame. In facts, two fuzzy systems with a similar structure are used corresponding to the vertical and horizontal motion components. The fuzzy system has two inputs (Input1, Input2) and one output. The fuzzy inputs are defined as:

$$\mathrm{x}_1 = \frac{1}{\mathrm{M}} \sum_{\mathrm{i}=\mathrm{n}-\mathrm{M}+1}^{\mathrm{n}} \left|\mathrm{GMV}_\mathrm{x}(\mathrm{i}) - \mathrm{GMV}_\mathrm{x}(\mathrm{i}-1)\right|, \quad (9)$$

$$\mathrm{x}_2 = \left|\mathrm{GMV}_\mathrm{x}(\mathrm{n}) - \mathrm{GMV}_\mathrm{x}(\mathrm{n}-\mathrm{M})\right|. \quad (10)$$

$$\mathrm{y}_1 = \frac{1}{\mathrm{M}} \sum_{\mathrm{i}=\mathrm{n}-\mathrm{M}+1}^{\mathrm{n}} \left|\mathrm{GMV}_\mathrm{y}(\mathrm{i}) - \mathrm{GMV}_\mathrm{y}(\mathrm{i}-1)\right|, \quad (11)$$

$$\mathrm{y}_2 = \left|\mathrm{GMV}_\mathrm{y}(\mathrm{n}) - \mathrm{GMV}_\mathrm{y}(\mathrm{n}-\mathrm{M})\right|. \quad (12)$$

where x_1 and x_2 denote the inputs of fuzzy system used for the adaptive filtering of the horizontal motion component and also y_1 and y_2 are the inputs of fuzzy system used for the adaptive filtering of the vertical motion component. $GMV_x(n)$ and $GMV_y(n)$ indicate the horizontal and vertical components of the GMV of last frame and M+1 is the number of last GMVs used for decision. The fuzzy system inputs, Input1 (x_1,y_1) and Input2 (x_2,y_2), are used as quantitative representations of unwanted and intentional camera movements, respectively. The value of Input1 is proportional to the noise amplitude and the value of Input2 is proportional to the intentional camera motion when it has an accelerating movement.

Defining suitable inputs for an adaptive DVS system has a great impact on the performance of system. Only relevant inputs can provide precise discriminating between unwanted and intentional camera motions to be used for the adaptation of Process Noise Variance. Different scenarios for the combination of unwanted and intentional camera motion can be considered. As examples, some scenarios are presented graphically in Fig. 5. In graphs (a) and (b), camera has an intentional accelerating movement plus noise or unwanted motion. The noise amplitude is high in (a) while it can be ignored in (b). Graph (e) is corresponding to a camera movement path while panning in which the camera is moving with a constant velocity without any acceleration and noise. The explanations of all graphs are summarized in Table2.

From the adaptive filtering point of view it is important to measure the amount of noise and the intentional camera movement velocity and acceleration. A softer Process Noise Variance is needed when the noise amplitude is high to remove the noise. On the other hand the soft Process Noise Variance prevents following of camera path when

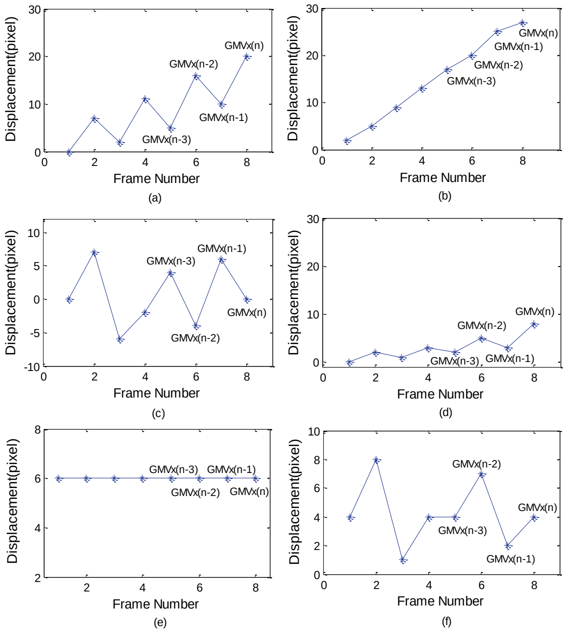


Fig.5 Sample scenarios for combination of unwanted and intentional camera motions: (a) high acceleration with high noise, (b) high acceleration with no noise, (c) high noise with no acceleration, (d) low acceleration with low noise, (e) constant velocity without noise and acceleration, (f) constant velocity with noise.

Table 2: Sample scenarios for combination of unwanted and intentional camera motion

Graph	*Noise*	*Velocity*	*Acceleration*
a	High	High	High
b	Low	High	High
c	High	Zero	Zero
d	Low	Low	Low
e	Zero	High	Zero
F	High	High	Zero

it has an intentional high acceleration. Therefore, the Process Noise Variance of Kalman filter should be tuned carefully proportional to the amount of noise and camera movement acceleration. According to this, we defined the fuzzy inputs so that Input1 gives information about the amount of noise and Input2 gives information about the amount of camera movement acceleration. It is notable that amount of camera movement velocity itself does not have any constrain on the filtering so it is not measured and used here. The proposed fuzzy system tunes the Process Noise Variance of the Kalman filter, Q(n), adaptively according to the amount of noise and the camera intentional accelerating movement. In the proposed fuzzy system, trapezoidal and triangular MFs are used for the inputs and the outputs, respectively. The number of MFs has been selected so as to obtain decent performance with as few MFs as possible to maintain low system complexity. The experimentally designed input and output membership functions and also the surface of desired outputs are shown in Fig. 6.

According to experimental results, the performance of used Kalman filter is more sensitive to Q's changes where Q has a small value. Therefore, more MFs of the fuzzy output are concentrated in this operating area. The constructed rule base is containing 30 rules as presented in Table3. The proposed fuzzy system was implemented while the *min* function was used for the *fuzzy implication* and the *max* function used for the *fuzzy aggregation*. Furthermore, the centroid defuzzification method was applied. The output of fuzzy system defines the Q of Kalman filter, i.e., Q (n). for MC of n^{th} video frame.

2.4.2 Adaptive Fuzzy MFs

Study on a number of video sequences has shown that the range of fuzzy inputs (Input1, Input2) is very variable on different video contents. Therefore, fixed MFs for the inputs of fuzzy system cannot provide a good stabilization performance over all video contents. In order to have a good performance for the proposed DVS system over different video contents, it is proposed to adjust the MFs of fuzzy inputs adaptively to recently received video frames. The range of MFs for the fuzzy inputs, i.e.,

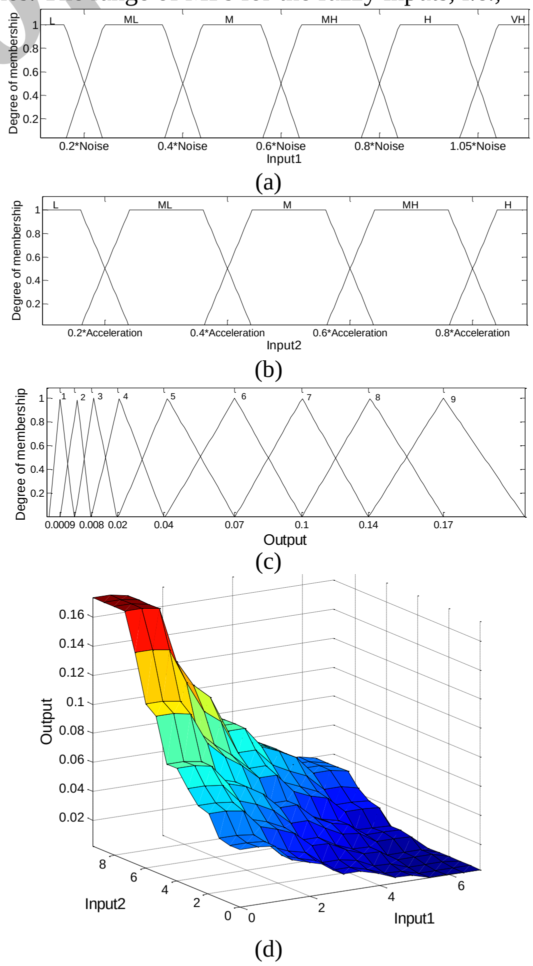


Fig. 6 (a) MFs of fuzzy Input1(x_1,y_1), (b) MFs of fuzzy Input2(x_2,y_2) (c) MFs of fuzzy output, (d) Surface of desired outputs.

(0, Input1(max)) and (0, Input2(max)) are modified adaptively as

Input1(max)= Max of input1over K recent frames, (13)

Input2(max)= Max of input2 over K recent frames, (14)

where Input1 and Input2 are clipped to a range from 1 to 10% of video frame height in term of pixel, and the K corresponds to the number of frames received in last few seconds, e.g., 2 s. This means that the system is adapted to the time-varying noise conditions while the frame size and frame rate are considered.

2.4.3 Motion Correction Algorithm

After computing the Q(n) (Process noise variance) by the fuzzy system, the camera motion path constructed by the

Table 3: Central Values of Fuzzy System Output*.

		Input1					
		L	ML	M	MH	H	VH
Input2	L	0.04	0.02	0.008	0.004	0.004	0.0009
	ML	0.07	0.04	0.02	0.008	0.004	0.0009
	M	0.1	0.07	0.04	0.02	0.008	0.0009
	MH	0.14	0.1	0.07	0.04	0.02	0.0009
	H	0.17	0.14	0.1	0.07	0.04	0.004

* L, low; ML, medium low; M, medium; MH, medium high; H, high; VH, very high

GMVs is filtered by the Kalman filter to compute the smoothened motion vectors (SMVs). For the first three frames, a fixed small value for Q(n) is used. After computing SMV, the unwanted motion vector (UMV) is obtained by

UMV (n) = GMV (n) - SMV (n). (15)

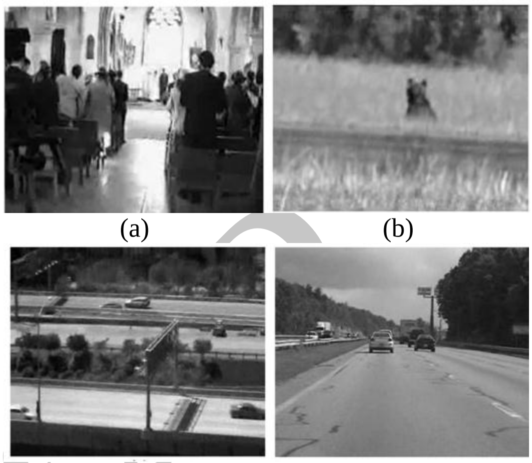
To restore the current frame to its stabilized position, we offset the current frame by the accumulated UMV, AMV, defined by

$AMV(n) = \sum_{i=m}^{n} UMV(i)$. (16)

Where *m* is the frame number of the last scene cut frame.

3. Experimental Results

The performance of the proposed DVS method is evaluated over 15 video sequences covering different types of scenes. Since there is no well-known video sequence in this research field, the algorithm is tested on a number of sequences which are easily available. For example, some used video sequences are available at [42,43]. These sequences have a frame rate of 25 fps and a picture size of 352×288 pixels. Sample frames of used video sequences are shown in Fig. 7. We worked with both gray-scale and color test sequences where in both cases ME is implemented on the luminance component. Good experimental results are obtained with M=3. However, a larger M provides more smoothness at the expense of more tracking delay and vice versa. Performance evaluation of the proposed DVS system are assessed according to the accuracy and computational complexity cost of the GMV estimation and also assessed according to the smoothness of the resultant global motion compared to the original sequence and the gross movement preservation capability.



(a) (b) (c) (d)

Fig. 7 Images taken by a camera (a-c) held by a hand; (d) in a moving vehicle

In this article, we propose a DVS algorithm with new features in ME and MC units. The ME unit estimates a GMV based on the BMVs which are estimated by the video encoder. Therefore, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. Moreover, in order to improve the accuracy of GMV estimation task an adaptive thresholding algorithm is used to remove the noisy invalid LMVs. It is notable that the adaptive thresholding algorithm uses MAD and MAD is computed during ME by encoder. Therefore, the adaptive thresholding algorithm have no additional computational complexity cost for the proposed DVS system. Regarding the mentioned modifications, we believe the ME unit of the proposed DVS outperforms the ME unit of the anchor systems, such as presented algorithms in [38,40], in terms of accuracy and computational cost. Therefore, only the performance of MC unit of proposed DVS system was compared with that of presented algorithms in [38,40] which are the most relevant known competitors. Fuzzy systems and adaptive Kalman filters are utilized for MC unit in [38,40]. The algorithms were applied on several data sets to simulate various scenarios. Since the performance of proposed algorithm in [38] depends on the value of Q parameter (*process noise variance*) of Kalman filter, so it was tuned with two different values of noise variance including Q=0.1 and Q=0.01. Some graphical comparison results are presented in Fig. 8. According to the results, the competitor algorithm proposed in [38] with small Q provides expand stabilization but reduces the capability of close tracking the intentional camera motion. Moreover, with a large Q, it enables close tracking of intentional camera movements but at the cost of slightly reduced stabilization capabilities. On the other hand,

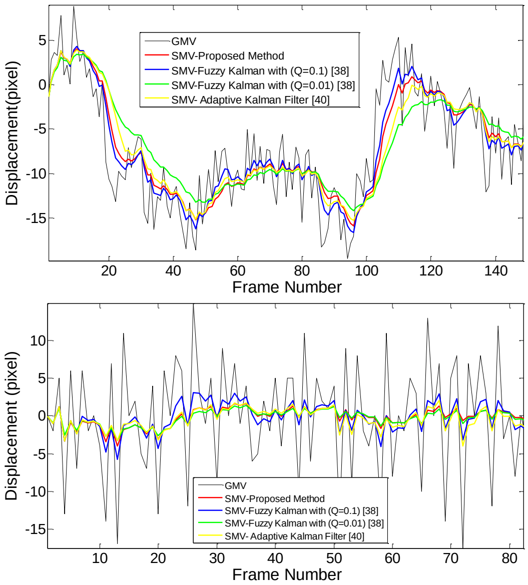


Fig. 8 Comparison results of proposed DVS algorithm with presented algorithm in [38] by two different Kalman filters: Q=0.1, Q=0.01, and presented algorithm in [40].

according to the comparison results, the proposed algorithm in [40] provides expand stabilization but reduces the capability of close tracking the intentional camera motion. Whereas results demonstrate that this article provides good stabilization and close tracking of intentional camera movements at the same time for all cases by adaptive tuning of the noise variance value. Numerical performance assessment of MC unit of a DVS system is a difficult task since the ground truth of unwanted motions as reference is not available. To solve this problem, we produced a synthetic signal as reference. The reference signal includes three parts corresponding to intentional camera movements with positive, negative, and zero value accelerations. Moreover, a sequence of normal random numbers was generated and the random numbers were rounded to integer values to simulate the unwanted camera motion in terms of pixel displacement. Provided sequence was added to the reference signal to obtain a synthetic signal including both the intentional and the unwanted camera motions. Two copies of the synthetic signal (y_1,y_2) were generated to be processed by MC units of the proposed DVS algorithm and also by the anchor algorithms presented in [38,40]. While the reference signal is known that the performance of compared algorithms can be evaluated numerically by computing a distance measure such as mean square error (MSE) between the reference signal as GMVs of the reference and the processed signals as smoothed GMVs. The MSE measure is computed between the smoothed GMVs and

the GMVs as:

$$MSE_{MV} = \frac{1}{N}\sum_{n=0}^{N}[GMV_R(n) - SMV(n)].^2 \qquad (17)$$

Where $GMV_R(n)$ and SMV(n) denote n^{th} GMV of the reference and the smoothed video sequence, respectively. It is noted the performance of proposed algorithm in [38] depends on the value of Q parameter (*process noise variance*) of Kalman filter, so it cannot provide a well daptation. Therefore, the two copies of the synthetic signal were processed by two values of Q (0.005 and 0.6) and graphical simulation results are presented in Fig.9 and Fig. 10, respectively. Moreover, numerical comparison results in term of MSE for the synthetic signals are presented in table 4.

According to the graphical results as shown in Fig.9 and Fig.10, the proposed algorithm in [38], depending on the Q value, performs only good filtering of the unwanted camera motion or only the tracking of intentional camera movement. On the other hand, the proposed algorithm

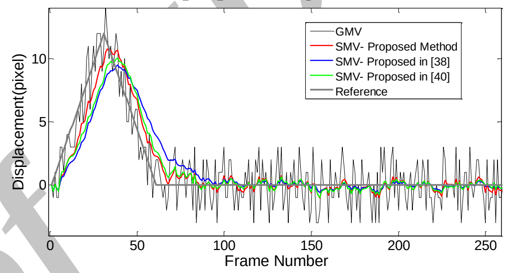


Fig. 9 Synthetic camera movement path y_1 processed by MC units of different DVSs

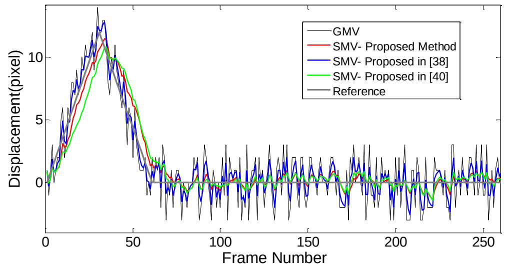


Fig. 10 Synthetic camera movement path y_2 processed by MC units of different DVSs

Table 4: MSE resulted by MC units of different DVSs

Motions Signal	*Adaptive Fuzzy Kalman Filter* (proposed)	*Fuzzy Kalman Filter, [38]*	*Adaptive Kalman Filter, [40]*
y_1	**0.9941**	**2.3239**	**1.5509**
y_2	**0.6045**	**1.0499**	**1.0579**

in [40] provides expand stabilization but reduces the capability of close tracking the intentional camera motion, respectively. Whereas the proposed algorithm in this article shows a high performance in both the removing of unwanted motion and the tracking of intentional camera movement. Furthermore, the numerical results presented in Table4 confirm the graphical results shown in Fig. 9 and Fig.10. The least MSE for our proposed algorithm means the best performance. A small-scale subjective quality test also demonstrated that human eyes have better visual perception to the stabilized videos by the proposed DVS system than the original videos in all cases.

4. Conclusions

In this article, we proposed a computationally efficient DVS algorithm using motion information obtained from a hybrid block-based video encoder. Since some of the obtained MVs are not valid, an adaptive thresholding was developed to filter out valid MVs and to compute an accurate GMV for each frame. The GMVs are smoothened with a kalman filter that is tuned adaptively to unwanted and intentional camera movements. The filter is adjusted by a fuzzy system with two inputs which quantify the unwanted and intentional camera movements. The proposed method fulfills two apparently conflicting requirements: close follow-up of the intentional camera movement and removal of the unwanted camera motion. In order to improve the stabilization performance, inputs MFs of the fuzzy system are continuously adapted according to the motion properties of a number of recently received video frames. Simulation results show a high performance for the proposed algorithm. With a low degree of computational complexity, the proposed scheme can effectively be used for the mobile video communications as well as for the conventional video coding applications to improve the visual quality of digital video and to provide a higher compression performance.

References

[1] A. Engelsberg and G. Schmidt,"A comparative review of digital image stabilizing algorithms for mobile video communications", IEEE Trans. Consumer Electron,Vol.45, No.3, 1999, pp. 592-597.

[2] Y.-C. Peng, C.-K. Liang, H.-A. Chang, H. H. Chen, and C.-J. Kao, "Integration of image stabilizer with video codec for digital video cameras", in *Proc. Int. Symp. Circuits Syst.*, 2005, Vol.5, pp.4781–4784.

[3] M. Okade and P.K. Biswas, "Fast Video Stabilization in the Compressed Domain", *IEEE International Conference on Multimedia and Expo (ICME),* 2012, pp. 1015-1020.

[4] H. H. Chen, C.-K. Liang, Y.-C. Peng and H.-A. Chang, "Integration of image stabilizer with video codec for digital video cameras", IEEE Trans. Video Technology, Vol. 17, No.7, 2007, pp. 801-813.

[5] L. Marcenaro, G. Vernazza and C. S. Regazzoni, "Image stabilization algorithms for video surveillance applications", *Proc. Int. Conf. on Image Processing.*, 2001, Vol.1, pp. 349-352.

[6] J. Zhou, H. He and D. Wan, "Video Stabilization and Completion Using Two Cameras", IEEE Transactions on circuit and systems for video technology, Vol. PP, No.99, 2011, pp. 1-1.

[7] M. Oshima, T. Hayashi, S. Fujioka, T. Inaji, H. Mitani, J. Kajino, K. Ikeda and K. Komoda, "VHS camcorder with electronic image stabilizer", *IEEE Trans. Consum. Electron*, Vol.35, No.4, 1989, pp. 749–758.

[8] K. Sato, S. Ishizuka, A. Nikami and M. Sato, "Control techniques for optical image stabilizing system", *IEEE Trans. Consum. Electron*,Vol. 39, No.3, 1993, pp. 461–466.

[9] S.-J. Ko, S.-H. Lee, S.-W. Jeon and E.-S. Kang, "Fast digital image stabilizer based on gray-coded bit-plane matching", *IEEE Transactions on Consumer Electronics*,Vol. 45, No.3, 1999, pp. 598– 603.

[10] A. Ertürk and S. Ertürk, "Two-Bit Transform for Binary Block Motion Estimation", IEEE Trans. Circuit Syst. Video Technol, Vol.15, No.7, 2005, pp. 938- 946.

[11] S. Ertürk, "Multiplication-free one-bit transform for low complexity block-based motion estimation", IEEE Sig. Proc. Letters, Vol.14, No.2, 2007, pp. 109-112.

[12] N.-J. Kim, H.-J. Lee and J.-B. Lee, "Probabilistic Global Motion Estimation Based on Laplacian Two-Bit Plane Matching for Fast Digital Image Stabilization", EURASIP Journal on Advances in Signal Processing, 2008.

[13] W.Nan, H. Xiaowei, W. Gang and Y. Zhonghu, "An approach of Electronic Stabilization Based on Binary image matching of Color Weight", 2nd International Asia Conference on Informatics in Control, Automation and robotics (CAR), 2010, Vol.2, pp. 155-158,.

[14] S. Battiato and G.Puglisi, "Fast block based local motion estimation for video stabilization", IEEE Computer Society Conference on Pattern Recognition Workshops (CVPRW), 2011, pp. 50-57.

[15] A.A. Amanatiadis, and I. Andreadis, "Digital Image Stabilization by Independent Component Analysis", IEEE. Instrumentation and Measurement, Vol.59, No.7, 2011, pp.1755-1763.

[16] A. A. Yeni and S. Ertürk, "Fast digital image stabilization using one bit transform based sub-image motion estimation", *IEEE Transactions on Consumer Electronics*, Vol.51, No.3, 2005, pp. 917– 921.

[17] S. Ertürk, "Digital Image Stabilization with Sub-Image Phase Correlation Based Global Motion Estimation", *IEEE Trans. Consumer Electron*, Vol.49, No.4, 2003, pp. 1320-1325.

[18] S. Ertürk, "Image sequence stabilization: Motion vector integration (MVI) versus frame position smoothing (FPS)", Proc. 2nd IEEE R8-EURASIP Symposium on Image and Signal Processing and Analysis, ISPA'01, 2011, pp. 266-271.

[19] T.-Han. Tsai, C.-Lun. Fang and H.-M. Chuang, "Design and Implementation of efficient Video Stabilization Engine Using Maximum a posteriori Estimation and Motion Energy Smoothing Approach", IEEE Transaction On Circuit and Systems For Video Technology, Vol.22, No.6, 2012, pp. 817-830.

[20] J. Zhu and B. Guo, "Fast Layered Bit-Plane Matching for Electronic Video Stabilization", International Conference on Multimedia and Signal Processing (CMSP), 2011, Vol.1, pp. 276-280.

[21] L. Dong-bok, C. Ick-hyun, S. Byung Cheol and L. Tae Hwan, "ROI-Based Video Stabilization Algorithm for Hand-Held Cameras", *IEEE International Conference on Multimedia and Expo Workshops(ICMEW)*, 2012, pp. 314-318.

[22] S. Erturk, "Translation, rotation and scale stabilization of image sequences", *IEE Electronics Letters*,Vol. 39, No.17, 2003, pp.1245-12462.

[23] N.A. Tsoligkas, S. Xalkiadis, X. Donglai and I. French, "A guide to digital image stabilization procedure- An overview", 18th International Conference on Systems signals and Image Processing (IWSSIP), 2011, pp. 1-4.

[24] J.M. Wang, H.P. Chou, S.W. Chen and C.S. Fuh, "Video stabilization for a hand-held camera based on 3D Motion model", 16th IEEE Internat*ional conference on Image Processing (ICIP), 2009*, pp. 3477-3480.
[25] O. Nestares, Y. Gat, H. Haussecker and I. Kozinsev, "Video stabilization to a global 3D frame of reference by fusing orientation sensor and image alignment data", 9th IEEE International symposium on Mixed and Augmented Reality (ISMAR), 2010, pp. 257-258.
[26] M. Mohammadi, M. Fathi and M. Soryani, "A new decoder side video stabilization using particle filter", 18th International Conference on Systems signals and Image Processing (IWSSIP), 2011, pp. 1-4.
[27] S.-H. Yang and F.-M. Jheng, "An adaptive image stabilization technique", In Proceedings of the *IEEE International Conference on Systems, Man, and Cybernetics* (SMC2006), 2006, Vol.3, pp. 1968-1973.
[28] F. Vella, A. Castorina, M. Mancuso and G. Messina, "Digital image stabilization by adaptive block motion vectors filtering", IEEE Trans. Consumer Electron, Vol.48, No.3, 2002, pp. 796-801.
[29] S. Battiato, G. Puglisi and A.R. Bruna, "A Robust Video Stabilization system by adaptive motion vectors filtering", IEEE International Conference on Multimedia and Expo, 2008, pp. 373-376.
[30] J.-P. Hsiao, C.-C. Hsu, T.-C. Shih, P.-L. Hsu, S.-S. Yeh and B.-C. Wang, "The Real-time Video Stabilization for the rescue Robot", ICROS-SICE International Joint conference, 2009, pp. 4364-4369.
[31] K. Uomori, A. Morimura and H. Ishii, "Electronic image stabilization system for *video* cameras and VCRs", *J. Soc. Motion Pict. Telev. Eng.*, Vol.101, 1992, pp. 66-75.
[32] S. Ertürk and T.J. Dennis, "Image sequence stabilization based on DFT filtering", *IEE Proc. on Image Vision and Signal Processing*, Vol.127, No.2, 2000, pp. 95-102.
[33] S. Ertürk, "Real-time digital image stabilization using Kalman filters", *Real-Time Imaging*, Vol.8, 2002, pp. 317-328.
[34] M.K. Güllü, E. Yaman and S. Ertürk, "Image sequence stabilization using fuzzy adaptive Kalman filtering", *Electronics Letters*, Vol.39, No.5, 2003, pp. 429-431.
[35] M.K. Güllü and S. Ertürk, "Fuzzy Image Sequence Stabilization", *Electronics Letters*, Vol. 39, No.16, 2003, pp. 1170-1172.
[36] M.K. Güllü and S. Ertürk, "Image sequence stabilization using membership selective fuzzy filtering", *Lecture Notes on Computer Science (LNCS)*, Vol.2869, 2003, pp. 497-504.
[37] M.K. Güllü and S. Ertürk, "Membership Function Adaptive Fuzzy Filter for Image Sequence Stabilization", *IEEE Trans. Consumer Electronics*, Vol.50, No.1, 2004, pp. 1-7.
[38] N. Kyriakoulis and A.Gasteratos, "A Recursive Fuzzy System for Efficient Digital Image Stabilization", *Hindawi Advances in Fuzzy Systems*, 2008.
[39] E. Yaman and S. Erturk, "Image Stabilization by Kalman Filtering Using a Constant Velocity Camera Model with Adaptive Process Noise", Proc. Of ELECO'2001, 2001, pp.152-155
[40] C. Wang, J-H. Kim, K-Y. Byun, J. Ni and S-J. Ko, "Robust Digital Image Stabilization Using the Kalman Filter", IEEE Transaction on Consumer Electronics, Vol.55, No.1, 2009, pp.6-14.
[41] B. Pinto and P.R. Anurenjan, "Video stabilization using speeded Up Robust features", International conference on communications and Signal processing (ICCSP), 2011, pp. 527-531.
[42] Road, http://www.jnack.com/adobe/photoshop/videostabilization/
[43] Shaky Car, matlabroot\toolbox\vipblks\vipdemos\shaky_car.avi

Mohamad Javad Tanakian was born in Zabol, Sistan and baluchestan, Iran in 1985. He received the B.S and also M.Sc degrees in Electrical engineering in 2007 and 2011 from university of Sistan and Baluchestan. His research interests are in the area of signal processing, Image and video processing and Fuzzy systems.

Mehdi Rezaei (IEEE M'04) received his B.Sc. degree in electronics engineering from Amir Kabir University of Technology (Polytechnic of Tehran) in 1992, his M.Sc. degree in electronics engineering from Tarbiat Modares University of Tehran in 1996, and his PhD degree in Signal Processing from Tampere University of Technology, Finland, in 2008. Dr. Rezaei had a closed cooperation with the Nokia Research Center as researcher and senior researcher from 2003 to 2009.
Dr. Rezaei is an Assistant Professor and also Deputy Dean for Education and Postgraduate Studies, in Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Iran. His research interests include multimedia signal processing and communications. He has published several journal and conference papers in these fields. He also holds several patents in related applications. Dr. Rezaei was the recipient of the Nokia Foundation Award in 2005 and also in 2006. He is a member of IEEE SP, COMSOC, and CAS Societies. Moreover, he is also a member of the Iranian computer vision and image processing society and also Iranian fuzzy systems society.

Farahnaz Mohanna received her B.Sc. degree in electronics engineering from University of Sistan and Baluchestan, Iran in 1987, her M.Sc. degree in electronics engineering from Tehran University in 1992, and her Ph.D. degree in Image Processing from Surrey University, Guilford, UK, in 2002. Dr. Mohanna has been working as a research fellow at the Centre for Vision, Speech and Signal Processing (CVSSP) at the Surrey University, UK in 2003. Dr. Mohanna is an Assistant Professor and Faculty Dean of Electrical and Computer Engineering, University of Sistan and Baluchestan, Iran. Her research interests include corner detection, active contours, object tracking, and video database retrieval and communications. She has published several Journal and conference papers in these fields. She is a member of the computer vision and image processing society.