Image sequence stabilisation: Motion vector integration (MVI) versus frame position smoothing (FPS)

Sarp ERTÜRK
Department of Electronics and Telecommunications Engineering,
University of Kocaeli, Izmit 41100, Turkey
s_erturk@hotmail.com

Abstract

Image sequence stabilisation is the task of removing unwanted camera movements (jitter) from an image sequence. While these jitter can be translational, rotational or zoom based, translational cases have larger influence on visual perception and are most likely to result in visual appearance degradation that will upset the viewer. While most of the research in the area of image stabilisation has been concentrated in the motion estimation part with the goal of accurate global motion estimation, the correction part has been given less attention.

Two techniques have been reported for the correction of translational jitter: motion vector integration (MVI) and frame position smoothing (FPS). While most reported stabilisation systems have employed MVI, this paper shows that FPS outperform this technique in that jitter stabilisation is accomplished while successfully preserving intentional camera motions, without the need to compromise as it is the case in MVI.

1. Introduction

Image sequence stabilisation (ISS) is the process of removing the effects of unwanted camera movements (jitter) from an image sequence. Undesired positional fluctuations, rotation or zoom are eliminated to produce a compensated image sequence that displays requisite camera movements only [1-17]. The discomfort experienced by the viewer due to jiggle and shake contained in video is thus eliminated; however intentional camera movements such as camera pan need to be retained in order to maintain video content. Therefore ISS systems are urged not to eliminate the entire global motion detected in the sequence, but need to resolve the jitter component so as to preserve intentional movements.

The image stabilisation process can be separated into two parts: 1) the motion estimation system 2) the motion correction system. ISS systems can accordingly be classified into the following categories:

i. Mechanical-Optical systems: Vibration feedback is achieved via gyro sensors [2], and stabilisation is accomplished by adjusting the refraction angle [3].

ii. Mechanical-Digital systems: Absolute vertical of the camera is tracked by a sensor [4] and movement correction is accomplished by digital processing.

iii. Fully Digital systems: Global motion estimation and movement correction are both performed by digital processing units [5-17]. Presently all utilised ISS systems are fully digital, as this enables inexpensive and lightweight implementation.

Present ISS schemes can further be categorised according to the global motion objective. Two-dimensional schemes compensate for translational jitter only, while three-dimensional schemes additionally compensate for rotation and scale variation.

Most of the research in the field of digital image stabilisation (DIS) is concentrated in the motion estimation part of the system. Two-dimensional ISS schemes require the estimation of translational motion only [5-12]. Three-dimensional ISS schemes require the estimation of rotation and scale parameters in addition to translation [13-15]. In [5] global motion estimation (GME) by processing local motion vectors of subimages obtained by block-matching has been proposed. GME based on edge pattern matching has been demonstrated in [6]. A multiresolution iterative motion estimation scheme that estimates affine motion parameters between levels of the Laplacian pyramid has been described in [7]. In [8] a pyramidal hardware system that uses mosaic-based registration has been presented. Morimoto and Chellappa have described a multiresolution feature-based motion estimation system in [9]. Motion estimation using bit-plane matching and gray-coded bit-plane matching has been demonstrated in [10] and [11] respectively. Ertürk and Dennis have presented an ISS system that makes use of phase correlation based motion estimation in [12]. In [13] a technique to compensate for rotation by tracking visual cues, with 3-D motion parameters estimated via Kalman filtering has been presented. Duric and Rosenfeld have used an image flow algorithm to estimate global motion [14]. Irani et al. have presented a technique that initially estimates 2-D affine motion followed by the computation of 3-D translation parameters of the camera [15].

In some two-dimensional ISS schemes proposed [7-9] the entire estimated motion is removed from the sequence not preserving intentional camera movements. Motion vector integration (MVI) [5,10,11,16] and frame position smoothing (FPS) [12] have been proposed as techniques to remove translational jitter with the aim to preserve intentional motion components. Three dimensional stabilisation systems are mainly aimed at removing [13-15] or smoothing [17] the rotational motion, either with stabilised or unprocessed translational component according to the desired application.

ISS systems are commonly integrated into consumer cameras and camcorder products to remove jitter and provide comfortable viewing of image sequences. Furthermore 3G mobile products are currently in the development stage, and as video communications will be a key element, it can be expected that ISS systems will be integrated to remove jitter resulting from vibrations encountered during operation. In a like manner for cases where video is captured from a unsteady camera, e.g. a camera mounted on a vehicle progressing on an unsteady surface or even on a moving helicopter (as it is sometimes the case in video taken for broadcast-news), image vibrations need to be removed to improve visual quality. In these cases two-dimensional stabilisation makes up the core of the ISS system, as the resulting translational jitter is most likely to result in visual appearance degradation that will disturb the viewer.

This paper compares motion vector integration (MVI) and frame position smoothing (FPS) techniques proposed for translational jitter stabilisation. These two techniques are primarily investigated according to their ability in jitter stabilisation and preservation of intentional movements. No concern is given on the motion estimation process in this paper: it is assumed that global interframe motion vectors have already been acquired with one of the available methods.

2. Translational jitter stabilisation

While image sequence stabilisation aims to remove undesired global motions from a video sequence and preserve intentional movements, it is required to discriminate between desired and undesired motion components. This judgement can be performed on the basis of smoothness: desired motions such as camera pan represent exclusively regular movements, while jitter is random of nature. In the frequency spectrum regular movements comprise lower frequencies, while jitter constitutes the higher frequency range, enabling the system to perform a frequency based decision. Hence low-pass filtering of the estimated global motion vectors, will thus retain intentional motion components and at the same time eliminate jitter.

Subsequent to the separation of the intentional motion component, stabilisation is achieved by shifting each image frame by an appropriate amount (namely the correction vector) so as to bring the frame into the position at which only desired global motion is contained. A comparative review of the two techniques, FPS and MVI, proposed to accomplish the task of resolving appropriate correction vectors follows.

2.1. Motion vector integration (MVI)

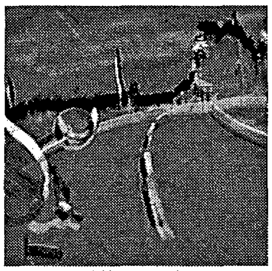
Motion vector integration yields correction vectors by integrating the differential global motion vector of a frame using a damping coefficient:

$$V_{\text{int}}(n) = k \times V_{\text{int}}(n-1) + V_{act}(n) \qquad (1)$$

The sampling index n is used to represent the frame number of the present frame in the image sequence. The damping parameter is shown by k: in practice, the damping factor value is chosen to be between 0.875 and 0.995 according to the required stabilisation intensity [5]. The notation $V_{act}(n)$ is used for the actual estimated global motion vector of the current frame with respect to the preceding frame. $V_{int}(n)$ is the integrated motion vector computed for the current frame. $V_{int}(n-1)$ is the integrated motion vector computed for the previous frame. The integrated motion vectors calculated for the present frames are used directly as correction vectors to translate image frames into their stabilised position. Thus the MVI system constitutes a first order IIR filter, that directly low-pass filters the differential motion vectors to remove interframe jitter.

Figure 1 (a) and (b) show sample frames from the bike sequence captured using a VHS camcorder mounted rigidly to one side of the rear carrier of a motorcycle, aiming past the rider. The sequence was passed through a time-base corrector that reduces intrafield jitter without affecting global interframe jitter. The difference image shown in Figure 1(c) exposes the comprehensive interframe jitter caused by camera vibration.

(a) Frame # 40 (b) Frame # 41



(c) Difference image

Figure 1: Sample frames for the bike sequence

Figure 2 shows the motion vectors that have been computed for the bike sequence (motion vectors in pixels vs. frame number). The motion vector oscillations clearly display position vibrations in the image sequence.

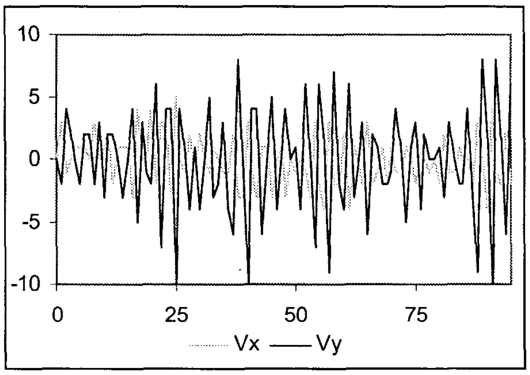
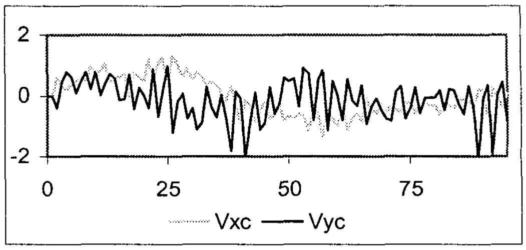
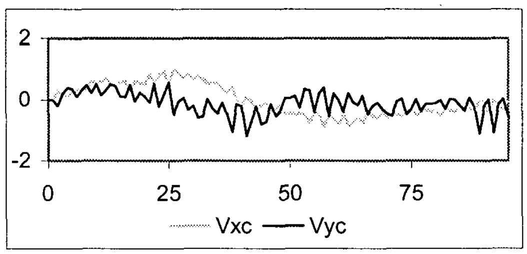


Figure 2: Computed motion vectors for the bike seq.

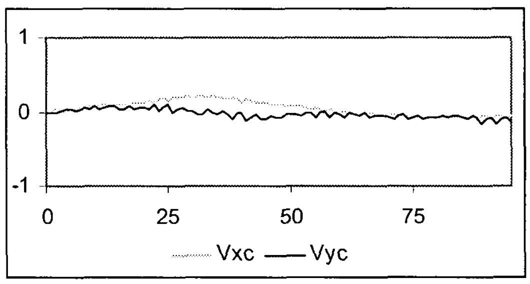
The motion vector of frame *n* represents the position shift of frame *n* with respect to the previos frame *n-1*: $V_{act}(n) = X_{initial}(n) - X_{initial}(n-1)$. For stabilisation the computed motion vectors are integrated using equation (1) and frames are shifted into their stabilised positions: $X_{corrected}(n-1) = X_{initial}(n-1) - V_{\text{int}}(n-1)$ and $X_{corrected}(n) = X_{initial}(n) - V_{\text{int}}(n)$. As the final motion vector for frame *n* after stabilisation is expressed in the form of $V_{corrected}(n) = X_{corrected}(n) - X_{corrected}(n-1)$, it is found that $V_{corrected}(n) = V_{act}(n) - V_{\text{int}}(n) + V_{\text{int}}(n-1)$. Figure 3 shows the resultant motion vectors for the MVI stabilised bike sequence for various damping parameters. It can be seen that motion vector amplitudes are significantly reduced, thus reducing jitter (**jitter is not entirely eliminated but reduced to subpixel amplitudes for high damping factors**). Stabilisation intensifies with increased damping coefficient, for a damping factor of k=0.99 the contained motion is nearly zero.



(a) k=0.8



(b) k=0.9



(c) k.=0.99

Figure 3: Motion vectors for MVI stabilised bike seq.

The global motion behavior of the image sequence becomes clearer by the investigation of absolute frame positions. Figure 4 shows the unprocessed absolute frame positions for the bike sequence, computed by accumulating differential motion vectors. Because of the left turn initially executed by the motorbike during filming, frames progress to the right hand side horizontally (in the positive X direction), then the subsequent right turn results in horizontal frame advances in the opposite direction. Horizontal jitter is contained throughout all parts of the sequence together with vertical jitter, resulting from camera vibrations.

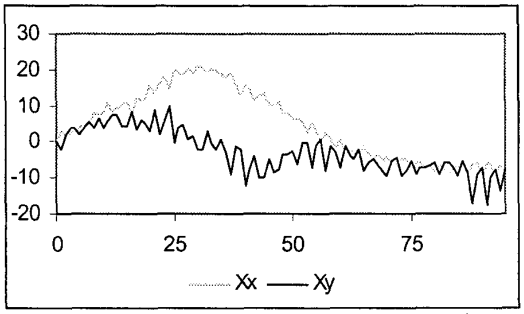
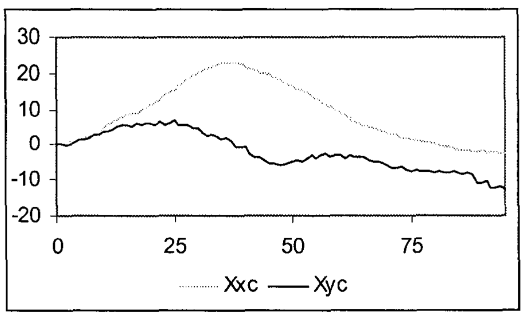
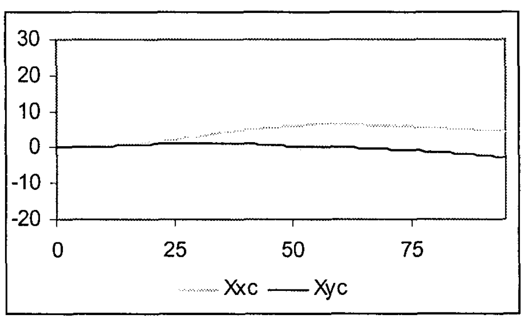


Figure 4: Absolute frame positions for bike seq.

Absolute frame positions for MVI stabilised sequences for damping factors of k=0.8 and k=0.99 are shown in Figure 5. It is seen in Figure 5 (b) that MVI can loose track of global sequence motions if the damping factor is too high, while stabilisation performance is reduced for too small damping factors as can be seen in Figure 5 (a). Thus the damping factor has to be selected to compromise stabilisation intensity and intentional motion preservation, for instance by using k=0.9, by the cost of loosing a bit of booth.



(a) k=0.8



(b) k=0.99

Figure 5: Absolute frame positions for MVI stabilised bike sequence

It has been proposed to select the damping parameter adaptively, so that a smaller damping coefficient is employed for sequences displaying rational motion (thus following global camera motions at the cost of low stabilisation intensity). While larger damping factors are engaged for sequences showing small global motions (concentrating on the stabilisation performance).

2.2. Frame position smoothing (FPS)

Frame position smoothing obtains correction vectors from the smoothed (lowpass filtered) absolute frame position signal that is constructed by accumulating global motion vectors. Acquired global interframe motion vectors are accumulated to yield an absolute frame position vs. frame number signal, say $X_{act}(n)$. This signal is lowpass filtered to remove high-frequency components resulting from translational jitter, and retain low frequency parts representing intentional camera displacements. Referring as $X_{lpf}(n)$ to the lowpass filtered absolute frame position, this represents the absolute position each frame should be brought into, with respect to the first frame, to have the sequence display desired global motions only. Therefore the correction vector for frame n is determined by

$$V_{cor}(n) = X_{lpf}(n) - X_{act}(n) \qquad (2)$$

Figure 6 shows original and low-pass filtered absolute frame positions for the bike sequence, having employed a Gaussian lowpass filter with a width of σ=6. Filtering is performed in the DFT domain by post-processing the absolute frame position signal. It is clearly seen that the filtered signal is a perfectly smoothed version of the original frame position signal.

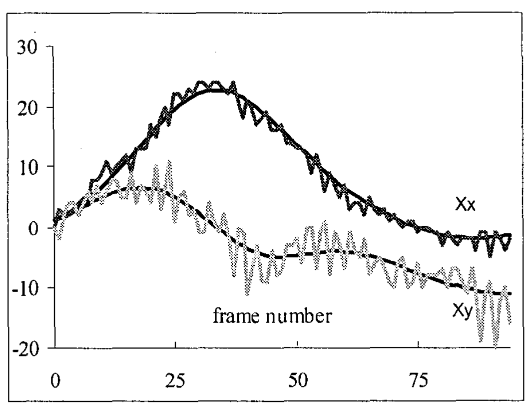
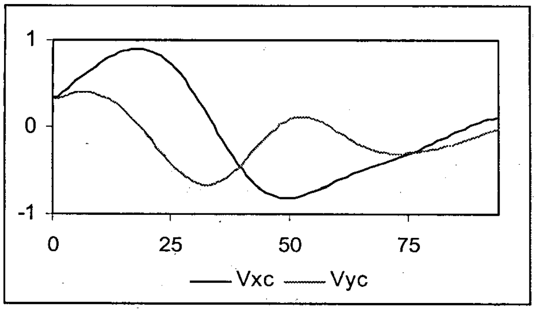


Figure 6: Original and DFT domain low-pass filtered absolute frame positions for the bike sequence

Figure 7 shows the resultant motion vectors for the FPS stabilised bike sequence. Compared to the MVI stabilised motion vectors given in Figure 3, the smooth global motion reconstruction of FPS is clearly noticeable.



(a) Gaussian low-pass, σ=6

Figure 7: Motion vectors for FPS stabilised bike seq.

However DFT domain filtering has the important drawback of requiring off-line post-processing of sequences, while MVI is effortlessly implemented for real-time operation. The utilisation of a digital filter largely overcomes this drawback of the DFT domain filter, as digital filters are characterised in terms of recurrence relations. For instance at an incoming picture rate of 25 Hz, a digital Butterworth low-pass IIR filter of order 4 with a corner frequency of 0.5 Hz is implemented by the following recurrence relation:

y[n] = (x[n- 4]+4x[n- 3]+6x[n- 2]+4x[n- 1]+x[n- 0]) + (-0.7199103273 * y[n- 4]) + (3.1159669252 * y[n- 3]) + (-5.0679983867 * y[n- 2]) + (3.6717290892 * y[n- 1])

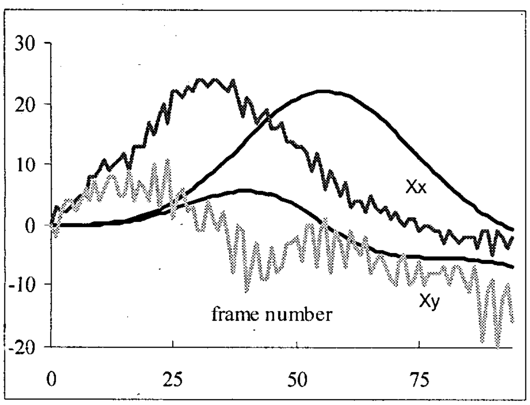


Figure 8: Original and IIR low-pass filtered absolute frame positions for the bike sequence

Figure 8 shows the original and IIR low-pass filtered absolute frame positions for the bike sequence. It can be seen from the Figure that the filter is successful in constructing a smooth absolute displacement signal, however the constructed signal is delayed with respect to the input. From the Butterworth filter impulse response that is shown in Figure 9, it is seen that the filter introduces a delay of about 25 samples (corresponding to 1 second). Thus for appropriate operation this 1 second delay has to be incorporated into the system. Figure 10 shows the IIR low-pass filtered displacements with the 1 second delay integrated so that the filtered output value $X_{lpf}(n+25)$ is assigned to correspond to the input displacement $X_{act}(n)$. Therefore the correction vector for frame n is obtained as $V_{cor}(n) = X_{lpf}(n+25) - X_{act}(n)$. IIR filtering thus enables nearly real-time operation of FPS, with an essential filter delay of one second incorporated.

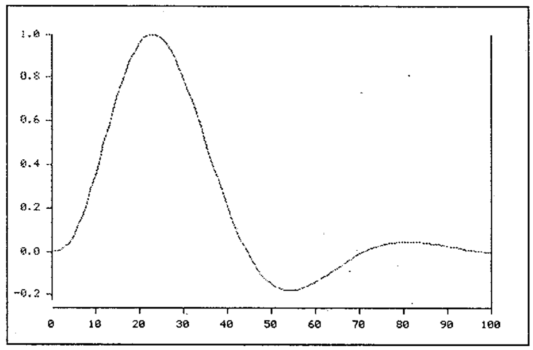


Figure 9: Impulse response of the Butterworth filter

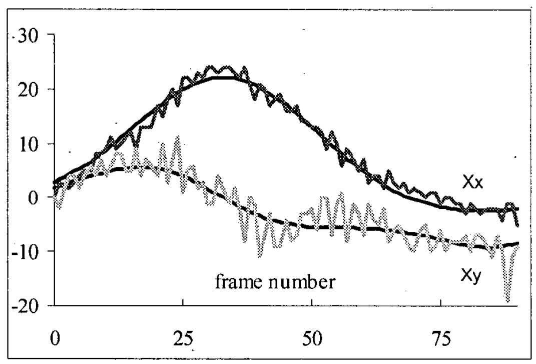


Figure 10: Original and IIR low-pass filtered (delayed by 1 second) absolute frame positions for the bike sequence

Currently implemented FPS systems operate either off-line (employing DFT filtering) or introduce a one-second filtering delay (employing IIR filtering). While off-line processing is feasible for some applications such as archive film restoration and non-live TV broadcast programs, in cases such as camcorder or live broadcast applications the one-second delay of FPS introduced by IIR filtering might be tolerated considering other delays caused for instance by processing and transmission.

3. Conclusions

Motion vector integration (MVI) and frame position smoothing (FPS) techniques proposed for translational jitter stabilisation have been compared in this paper. It is shown that MVI needs to compromise between stabilisation performance and intentional motion preservation, while FPS accomplishes a smooth reconstruction of the actual long-term camera motion by filtering out jitter components. It is observed from the final motion vectors of stabilised sequences that MVI only dampes the jitter amplitude below a certain level, while IPS obtains a perfectly smooth frame transition.

In MVI, stabilisation intensity is adjusted by varying the damping parameter value, however either stabilisation performance is reduced (by decreased damping) or intentional motion preservation performance is diminished (by increased damping). In the case of FPS, stabilisation intensity can be adjusted by setting filter characteristics (order of the IIR filter and cut-off frequency), at any case stabilisation and intentional motion preservation will be optimised.

MVI has the advantage of providing simple real-time implementation possibility. While currently realised FPS systems operate either off-line by DFT filtering or with a one-second delay introduced by IIR filtering of absolute frame position, future work is due to investigate alternative filtering techniques to achieve real-time FPS.

4. References

[1] A. Rosenfeld: "Guest Editorial", *Real-Time Imaging*, **2**, p. 269, 1996.

[2] M. Oshima, Y. Hayashi, S. Fujiko, T. Inaji, H. Mitani, J. Kajino, K. Ikeda, and K., Komoda: "VHS camcorder with electronic image stabilizer", *IEEE Trans. Consum. Electron.*, **35**, (4), pp. 749-757, 1989.

[3] K. Sato, S. Ishizuka, A. Nikami, and M. Sato: "Control techniques for optical image stabilizing system", *IEEE Trans. Consum. Electron.*, **39**, (3), pp. 461-465, 1993.

[4] T. Vieville, E. Clergue, and P.E.D.S. Facao: "Computation of ego-motion and structure from visual and internal sensors using the vertical cue", *Proceedings of IEEE Int. Conf. On Computer Vision*, Berlin, pp. 591-598, 1993.

[5] K. Uomori, A. Morimura, and H. Ishii: "Electronic image stabilization systems for video cameras and VCRs", *J. Soc. Motion Pict. Telev. Eng.*, **101**, (2), pp. 66-75, 1992.

[6] J.K. Pail, Y.C. Park, and D.W. Kim: "An adaptive motion decision system for digital image stabilizer based on edge pattern matching", *IEEE Trans. Consum. Electron.*, **38**, (3), pp. 607-615, 1992.

[7] P. Burt, and P. Anandan: "Image stabilization by registration to a reference mosaic", *Proc. of ARPA Image Understanding Workshop*, Monterey, CA, pp. 425-434, 1994.

[8] M. Hansen, P. Anandan, K. Dana, G. Van Der Wal, P.J. Burt: "Real-time scene stabilization and mosaic construction", *Proc. of ARPA Image Understanding Workshop*, Monterey, CA, pp. 457-465, 1994.

[9] C. Morimoto, and R. Chellappa: "Fast electronic digital image stabilization for off-road navigation", Real-Time Imaging, **2**, pp. 285-296, 1996.

[10] S.J. Ko, S.H. Lee, and K.H. Lee: "Digital image stabilizing algorithms based on bit-plane matching", *IEEE Trans. Consum. Electron.*, **44**, (3), pp. 617-622, 1998.

[11] 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 Trans. Consum. Electron.*, **45**, (3), pp. 598-603, 1999.

[12] S. Ertürk, and T.J. Dennis: "Image sequence stabilisation based on DFT filtering", *IEE Proc. On Vision Image and Signal Processing*, **147**, (2), pp. 95-102, 2000.

[13] Y.S. Yao, P. Burlina, and R. Chellappa: "Electronic image stabilization using multiple visual cues", *Proc. of IEEE Int. Conf. on Image Processing*, Washington DC, 1995

[14] Z. Duric and A. Rosenfeld: "Image sequence stabilization in real time", *Real-Time Imaging*, **2**, pp. 271-284, 1996.

[15] M. Irani, B. Rousso, and S. Peleg: "Recovery of ego-motion using image stabilization", *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Seattle, pp. 454-460, 1994.

[16] A. Engelsberg, and G. Schmidt: "A comparative review of digital image stabilising algorithms for mobile video communications", *IEEE Trans. Consum. Electron.*, **45**, (3), pp. 591-597, 1999.

[17] Y.S. Yao, and R. Chellappa: "Selective stabilization of images acquired by unmanned ground vehicles", *IEEE Trans. Robotics and Automation*, **13**, (5), pp. 693-708, 1997.