Efficient Video Stabilization Technique for Hand Held Mobile Videos

Paresh Rawat[1] and Jyoti Singhai[2]

*[1]Department of Electronics & Communication Engineering,
Truba College of Science &Technology Bhopal, India
[2]Department of Electronics Engineering
Maulana Azad National Institute of Technology, Bhopal, India*

[1]parrawat@gmail.com, [2]j.singhai@gmail.com

Abstract

Majority of the videos that have been captured by mobile cameras are suffering from low quality due to either low end manufacturing designs or complicated operating environments and untrained users. Thus videos taken by hand held mobile cameras tend to suffer from different undesired slow motions that cause annoying shaky motion and jitter. It is desirable to stabilize the video sequence by removing the undesired motion between the successive frames. Current methods are applicable to only specific camera motion models; hence having limitation to process gorse motion. In this paper an efficient video stabilization algorithm for hand held camera videos has been proposed. The proposed algorithm uses differential global motion estimation with Taylor series expansion to improve the estimation efficiency. Affine motion model has been assumed to define the inter-frame error between consecutive frames. Motion vectors have been estimated analytically by solving the derivatives of the inter-frame error. After motion estimation Gaussian kernel filtering has been used to smoothen out estimated motion parameters. Inverse rotation smoothening has been applied to remove the rotation effect from the smoothed transformation chain. This has led to reduced accumulation error and minimizes the missing image area significantly. The performance of the proposed algorithm has been tested on real time videos and compared with existing algorithm.

Keywords: *Video stabilization, differential motion estimation, Taylor series expansion, Gaussian Kernel filtering, motion smoothing*

1. Introduction

Inventions of hand-held devices, such as digital camcorders and cell phones with video capturing capabilities, have enabled everyday users to capture high-quality videos. The video imagery can be processed as a sequence of still images, where each frame is processed independently. However, the utilization of existing temporal redundancy by means of multiframe processing enables us to develop more effective algorithms, such as video stabilization. Hence video stabilization is becoming an indispensable technique in improving the design of these mobile cameras. Stabilization is a video processing technique to enhance the quality of input video [1, 7]. Stabilization is achieved by synthesizing a new stabilized video sequence; by estimating and removing the undesired inter frame motion between the successive frames. The video stabilization can either be achieved by hardware or post image processing approach. Hardware approach can be further classified as mechanical or optical stabilization. Mechanical stabilizer uses gyroscopic sensor to stabilize entire camera. Optical stabilization activates an optical system to adjust camera motion sensors. This approach is expensive and also has limitation to process different kind of motions simultaneously. In the image post processing algorithm, there are typically three major stages constituting a video stabilization process viz. camera motion estimation, motion smoothing or motion compensation, and mage warping. There are various algorithms proposed for stabilizing videos taken under different environment from different camera systems by modifying these three stages.

The development of video stabilization can be traced back by the work done in the field of motion estimation. Various algorithms have been proposed to reduce the computational complexity and to improve the accuracy of the motion estimation. The efficiency of the stabilization depends on the accuracy of the motion estimation and optical flow methods. Horn and Schunck (HS) [17] is widely used optical flow method. But it only computes the slow motion and provides the motion vectors in one direction only. The paper discusses the optical flow performance of original and modified HS method using 1D separable filter to find the temporal derivatives.

The global motion estimation can either be achieved by feature based approaches [2-5] or direct pixel based approaches [1, 7-9]. The feature-based approaches are although faster than direct pixel based approaches, but they are more prone to local effects and their efficiency depends upon the feature point selection [1]. Thus they have limited performance for unintentional motion. The direct pixel based approach makes optimal use of the information available in motion estimation and image alignment, since they measure the contribution of every pixel in the video frame. Matsushita *et al.*, [1], in 2006 proposed the direct pixel based full frame video stabilization approach using hierarchical differential motion estimation with Gauss Newton minimization. After motion estimation, motion inpainting is used to generate full frame video. This method gave good results in most videos; except in those cases when large portion of video frame is covered by a moving object, since this large motion makes the global motion estimation unstable. In this paper a modified video stabilization algorithm for hand held camera videos is proposed. The proposed algorithm uses Taylor series expansion instead of Gauss Newton minimization. Property of Taylor series is that it converges for each value of motion vectors and hence provides stable global motion estimation. After motion estimation Gaussian kernel filtering is used, only to smoothen out estimated motion parameters. This reduces the accumulation error and minimizes missing image area significantly.

In this paper existing algorithms used to stabilize the different type of video sequences are discussed in Section 2. In Section 3 proposed hierarchical differential motion estimation and Gaussian kernel filtering for motion smoothing are discussed. The results obtained with the proposed video stabilization algorithm are discussed in Section 4.

2. Reviews of Video Stabilization Algorithms

Video stabilization can be broadly classified as mechanical stabilization, optical stabilization and image post processing stabilization. Mechanical video stabilization systems based on vibration feedback via sensors like gyros accelerometers etc. have been developed in the early stage of camcorders [22]. Optical image stabilization, which has been developed after mechanical image stabilization, employs a prism or moveable lens assembly that variably adjusts the path length of the light as it travels through the camera's lens system [23]. Mechanical and optical stabilization methods are unsuitable for small camera modules embedded in mobile phones due to lack of compactness and the cost associated with it. Hence digital video stabilizers with lesser complexity and fast response are more suitable for stabilizing the hand held mobile camera video.

The digital video stabilization methods can be broadly classified as direct pixel based methods [1, 7, 9, 15], and feature based methods as [2-5]. The efficiency of the motion estimation technique depends on the optical flow method used. Two most widely used optical flow methods are Horn and Schunck (HS) [17] and Lucas Kanade [18]. The performance of the optical flow methods depends on the method used to find the temporal derivatives. Hany Farid [7] used the 1D separable kernel filters to find the temporal derivatives. The majority of today's methods strongly resemble the original formulation of HS which is a global method. They combine a data term that assumes constancy of some image property with a spatial term that models how the flow is expected to vary across the image. An objective function combining these two terms is then optimized. On the other hand Lucas Kanade method is a local optical flow method. But these methods were initially used for the slow motion videos.

Various feature based approaches are proposed for video stabilization. Chang *et al.*, [2] presented a feature tracking approach based on optical flow, considering the fixed grid of points in the video. But this approach was developed for a specific motion model [2]. Rong Hu, *et al.*, [3] in 2007 proposed an algorithm to estimate the global camera motion with SIFT features. These SIFT features have been proved to be affine invariant and used to remove the intentional camera motions. Junlan Yang *et al.*, [4] in 2009 used SIFT feature points and particle filtering framework to estimate the global motion between two frames. To estimate intentional motion from accumulative motion Kalman filter is used. Derek Pang *et al.*, [5] in 2010 proposed the video stabilization using Dual-Tree complex wavelet transform (DT-CWT). This method uses the relationship between the phase changes of DT-CWT and the shift invariant feature displacement in spatial domain to perform the motion estimation. Optimal Gaussian kernel filtering is used to smoothen out the motion jitters. This phase based method is immune to illumination changes between images, but this algorithm is computationally complex. R. Szeliski, [6] in 2006 presented a survey on image alignment to explain the various motion models, and also presented a good comparison of pixel based direct and feature based methods of motion estimation. The efficiency of the feature based methods depends upon the feature point's selection [6]. The features would often be distributed unevenly over the images, hence feature based methods may fail to match image pairs that should have been aligned. The feature based methods may have probability to get confused in regions that were either too textured or not textured enough.

Direct pixel based methods use each pixel in the frame to estimate the global motion. Hany Farid and J.B. Woodward in 1997 [7], modelled motion between video frames as a global affine transform and parameters are estimated by hierarchical differential motion algorithms. Temporal mean and median filters were applied to this stabilized video sequence for enhancing the video quality. But they have not implemented the motion smoothening or compensation algorithms. Olivier Adda, *et al.*, [8] in 2003 presented various motion estimation and compensation algorithms for video sequences. They suggested the uses of hierarchical motion estimation with gradient descent search to converge the parameters. But the method was slow and complex.

Matsushita *et al.*, [1], in 2006 proposed the direct pixel based full frame video stabilization method with motion inpainting. They achieved video stabilization by assuming an affine motion model between each pair of frame to represent the inter frame error between adjacent frames. Then an 'L' level Laplacian image pyramid is constructed and inter-frame error is estimated using hierarchical differential motion estimation which leads to enhanced accuracy, robustness and improved efficiency [15-16]. Estimation process involves SSD minimization with Gauss Newton minimization, which uses a first order expansion of the individual error quantities before squaring. The limitation of this method is that it strongly relies on the result of global motion estimation which may become unstable when a moving object covers large amount of image area [1]. For fast moving objects neighbouring frames will not be warped correctly, and there will be visible artefacts at the boundaries. The convergence ability of the Gauss Newton minimization is also limited.

Feng Liu *et al.*, [9] in 2009 proposed an algorithm of content preserving warps for 3D video stabilization for hand held cameras. The key insight of the work is that for the purposes of video stabilization, small shift in viewpoint can be faked by a carefully constructed content preserving warp, but result is not physically accurate. The major limitation of this approach compared to 2D video stabilization is that it first requires running structure from motion, and method is also more brittle and heavy weight. R. Szeliski [6] suggested that, for matching the sequential frames in a video direct pixel based methods can be used. Direct methods make optimal use of the information available in image alignment and provide a very accurate alignment results. It is because they use each pixel in the frame to estimate the global motion. However, the computational load is heavy and convergence range is also limited.

After estimation to smoothen the undesired camera motion in the global transformation chain, various approaches have been proposed [10-13]. Buehler *et al.*, [10] proposed Image based rendering algorithm to stabilize video sequence. The camera motion was estimated by non-metric algorithm, and then image-based rendering was applied to smoothed camera motion. Method performs well only with simple and slow camera motion videos and was unable to fitter motion models to complex motion as in the case of hand held camera videos. Litvin *et al.*, [11] applied the probabilistic methods using Kalman filter to smoothen camera motion. This method produced very accurate results in most of the cases, but it required tuning of camera motion model parameters to match with the type of camera motion in the video. Matsushita *et al.*, [1] developed an improved method called Motion inpainting for reconstructing undefined regions and to smoothen camera motion Gaussian kernel filtering was used effectively.

3. Proposed Video Stabilization Algorithm

In this paper an efficient video stabilization method is proposed for hand held mobile phone cameras. The proposed method uses the differential global motion estimation with the combination of the Gaussian kernel filtering for motion smoothing. The video sequences are captured from mobile phone camera in different environment.

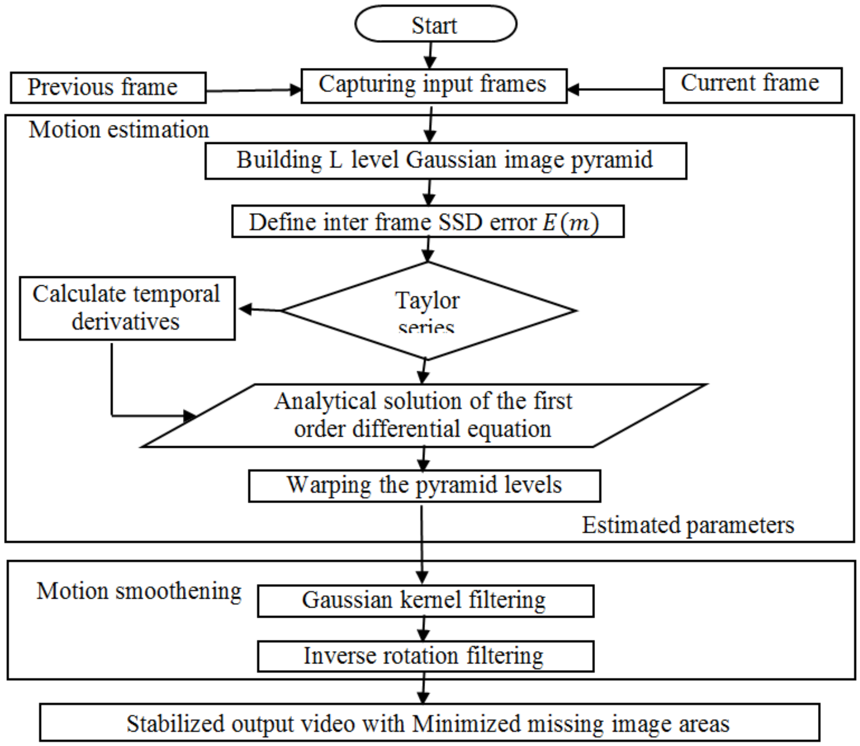


Figure 1. Proposed Algorithm for Hand Held Mobile Video Stabilization

The motion is first detected using the optical flow method and then motion vectors are estimated. The proposed algorithm improves the efficiency and convergence rate of the global motion estimation. This is achieved by using Taylor series expansion [7], instead of Gauss Newton minimization algorithm as proposed by Matsushita *et al.*, [1], for differential motion estimation. The use of Taylor series reduces the nonlinear error function to linear differential equation.

The linear differential equation can be solved analytically, hence reduces computational complexity. The proposed algorithm is explained in Figure 1, which consists of two stages motion estimation and motion smoothening. Motion estimation is explained in Section 3.1, in this stage every frame of video sequence is decomposed into L level Laplacian image pyramid. The motion between successive frames is estimated using first order Taylor series expansion. The temporal derivatives are determined by 1-D separable filters. In the presence of fast moving object in the frames, use of the bi-cubic interpolation for warping the pyramid levels minimizes the visible artefacts at the boundaries. In Section 3.2 motion smoothening is explained using Gaussian kernel filtering by smoothening estimated transform parameters to minimize the missing image areas. Then inverse rotation filtering is applied on smoothened frame to generate the window based completion method to reduce the overall accumulation error.

3.1. Motion Estimation

The motion of any pixel between two consecutive frames can be estimated either by global motion or by local motion. The global motion occurs due to camera motion but in local motion object in the scene is in motion. In case of a non-stationary camera or for small motion of the object, motion is estimated by a global motion model. The direct method of global motion estimation makes optimal use of the information available in image alignment, since they measure the contribution of every pixel in the video frame. For matching sequential frames in a video, the direct approach can usually be made to work [6]. The differential global motion estimation has proven highly effective at computing inter-frame motion [7, 14]. The method used in this paper for motion estimation is similar to that of the [7], the motion between two sequential frames, $f(x, y, t)$ and $f(x, y, t-1)$ is modelled with a 6-parameter affine transform as in [1, 7, 20, and 21]. The major advantage of using the affine model lies in the fact that for global motion, the affine parameters at every location should be the same. Therefore, instead of keeping track of every motion vector, the sum of square difference (SSD) error between two images can be described by a single affine transformation as given by eq. (1).

$$E(a) = \sum_{x,y \in \Omega} [f(x, y, t) - f(a_1 x + ay + a_5 , a_3 x + a_4 y + a_6 , t-1]^2 \tag{1}$$

Where $a_1, a_2,\ a_3,$ and a_4 represents the 2×2 affine rotation matrix A, and a_5 and a_6 the translation vector $\overline{T}$.

$$A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \ and \quad \overline{T} = \begin{pmatrix} a_5 \\ a_6 \end{pmatrix} \tag{2}$$

Where Ω denotes a user specified region of interest here it is the entire frame.

To improve the performance of the motion estimation a hierarchical global estimation is used [1]. An L -level Gaussian pyramid is built for each frame pair of frames $f(x, y, t)$and $f(x, y, t-1)$. The motion estimated at each pyramid level L is used to warp the frame at the next higher level $L-1$, until the finest level of the pyramid is reached (the full resolution frame is at $L = 1$). Large motions are estimated at coarse level by warping using bicubic interpolation and refining iteratively at each pyramid level. If the estimated motion at pyramid level L is $a_1, a_2, a_3, a_{4,}, a_5$ and a_6, then the original frame should be warped with the affine matrix A and the translation vector $\overline{T}$ is given by (3).

After working at each level of the pyramid, the original frame will have to be repeatedly warped according to the motion estimated at each pyramid level. Two affine matrices A_1 and A_2 and corresponding translation vectors $\overline{T}_1$ and $\overline{T}_2$ are combined as in equation (3), which is equivalent to applying A_1 and $\overline{T}_1$ followed by A_2 and $\overline{T}_{2.}$

$$A = A_1 A_2 \text{ , and } \overline{T} = A_2 \overline{T}_1 + \overline{T}_2 \tag{3}$$

To simplify the minimization, this error function (1) is approximated by using a first-order truncated Taylor series expansion as in method of Hany Farid [7]. The quadratic error function is now linear in its unknowns, *m* and can therefore be minimized analytically by differentiating with respect to *m* as,

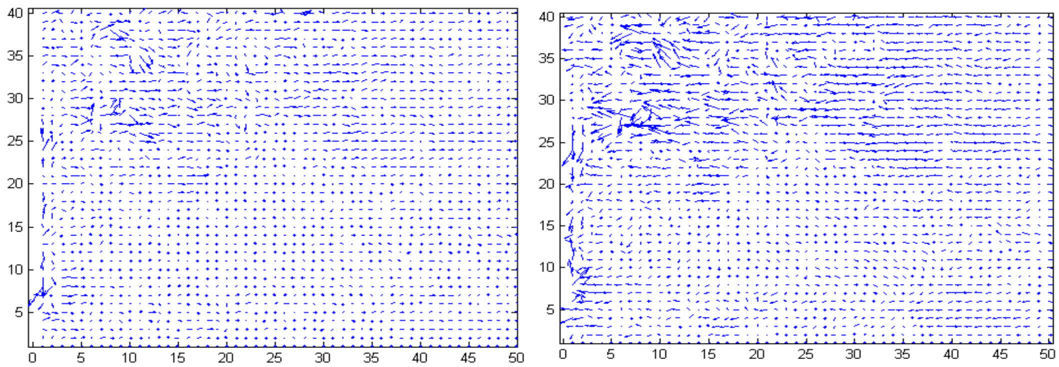
$$\frac{dE(a)}{da} = \sum_{\Omega} 2C[k - C^T a] \tag{4}$$

Where the scalar k and vector $C^T$are given as

$$k = f_t + x f_x + y f_y \ and \ C^T = \begin{pmatrix} x f_x & y f_x & x f_y & y f_y & f_x & f_y \end{pmatrix}$$

By setting the result equal to zero, and solving for *m* yields eq. (5).

$$a = \left[\sum_{\Omega} C\, C^T\right]^{-1} \left[\sum_{\Omega} C\, K\right] \tag{5}$$



a) With Horn and Schunck method(17] b) With proposed method

Figure 2. Comparison of the motion vectors for car video frame 15 and 16

The temporal derivatives are calculated by using 1-D separable kernel filters as in [7]. The advantage of using the separable kernel filter of $M\ x\ N$ is that the computation is reduced to $(M+N)$ multiplication from $(M\ x\ N)$ multiplication. Hence 1-D separable filters are used to reduce the computation complexity. The comparison of the optical flow of the original derivatives used by Horn and Schunck [17] and the derivatives with separable kernel are shown in Figure 2. Although estimating the global motion, but the motion vectors may be in different direction at different locations. It is clear that use of separable kernel for calculating the temporal derivatives performs better than conventional method because they are able to represent motion vectors more accurately.

3.2. Motion Smoothening

The proposed video stabilization algorithm uses Gaussian kernel filtering to smooth the undesired camera motion after motion estimation and to remove accumulation error. In order to avoid the accumulation error due to the cascade of original and smoothened transformation chain, displacement among the neighbor frames is smoothened to generate a compensation motion. The coordinate transformed from frame i to j, are denoted by the transform $\overline{T}_i^j$, as used by Matsushita *et al.* [1]. The neighbor frame is given as,

$$N_t = (m : t - k \leq m \leq t + k) \tag{6}$$

The idea of Gaussian smoothing is to use this 2-D distribution as a point spread function (PSF), and this is achieved by convolution.The compensation motion transform can be calculated as;

$$C_t = \sum_{i \in N_t} T_j^i * G(k) \tag{7}$$

Where $*$ means convolution operator and $G(k)$ is the Gaussian kernel given as

$$G(k) = \frac{e^{-k^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \quad (8)$$

The motion compensated frames I_t' can be warped from the original frame I_t as;

$$I_t' = C_t I_t \quad (9)$$

The use of large Gaussian kernel might lead to the blurring effects and small Gaussian kernel may not effectively remove the high frequency camera motion. Hence an optimal value of Gaussian kernel is selected. The parameter of Gaussian filter is set to as $\sigma = \sqrt{k}$ [1]. The σ value for Gaussian kernel should not be greater than 2.6. Hence the kernel parameter k should be either less than or equal to 6. Use of Gaussian kernel filtering minimizes the missing image areas.

3.2.1. Rotation Smoothing

The rotation effect caused due to the smoothed transformation chain is removed by rotating the frame by rotation angles. The rotation angles θ_1 and θ_2 are calculated by using the smoothed affine parameters m_2 and m_3 as

$$\theta_1 = sin^{-1}(m_2) \ and \ \theta_2 = sin^{-1}(m_3) \quad (10)$$

After calculating the rotation angles, the frame is rotated in reverse direction by these angles to remove rotation effects.

$$R = R(\theta_1) + R(\theta_2) \quad (11)$$

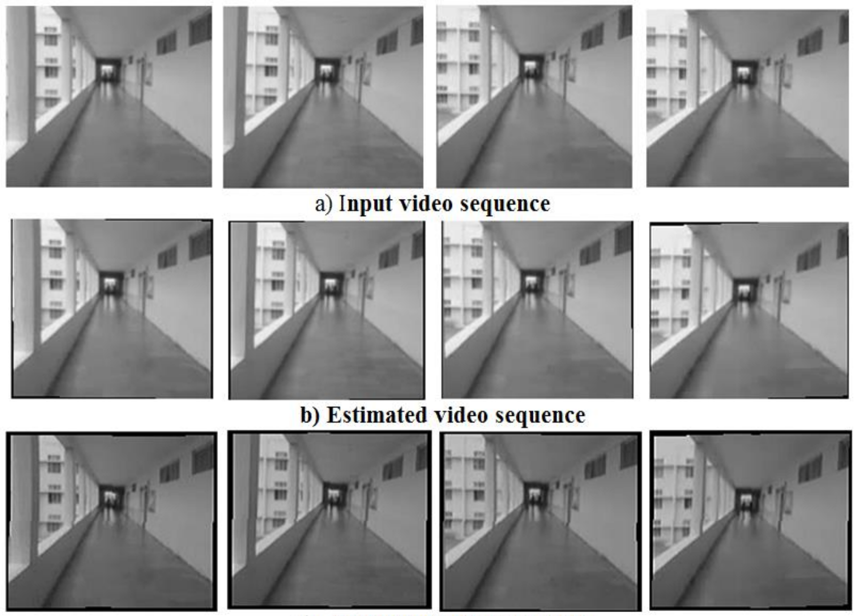
Where $R(\theta 1)$ and $R(\theta 2)$ are the inverse rotation factors. By using this inverse rotation factors the missing image areas are minimized.

4. Results and Discussion

Performance of the proposed video stabilization algorithm is tested on sixteen real time video sequences generated by Nokia (6303) mobile phone camera. The frame rate was 15 frames per second with the resolution of 176 x 144. The performance for two distinct videos viz. Corridor and Highway are illustrated in this paper for comparison with other algorithms. The Corridor video shown in Figure 3(a) is a slow motion video and Highway video shown in Figure 4(a) is a video with large object motion in static scene. The results of the optical flow methods of HS derivatives [17] and for 1D separable filters are presented for the case of large moving object in the scene. The difference of the original and estimated frames is used motion for the performance evaluation as in Figure 5. It is clear that proposed method is able to estimate more precise motion vectors and thus gave better results than basic HS method.

To stabilize the motion between each pair of frames in these videos global motion estimation has been used. To verify the performance of motion estimation, the inter frame error between original input frames were compared with, inter frame error after motion estimation with Mean filtering, Median filtering, Bicubic interpolation and Spline interpolation methods. The frame to frame comparison for MSE and SNR for original input and motion estimated video sequences are shown in Table 1 and Table 2 respectively.

Comparisons are shown for 10 consecutive frames having the maximum motion in the given two videos.

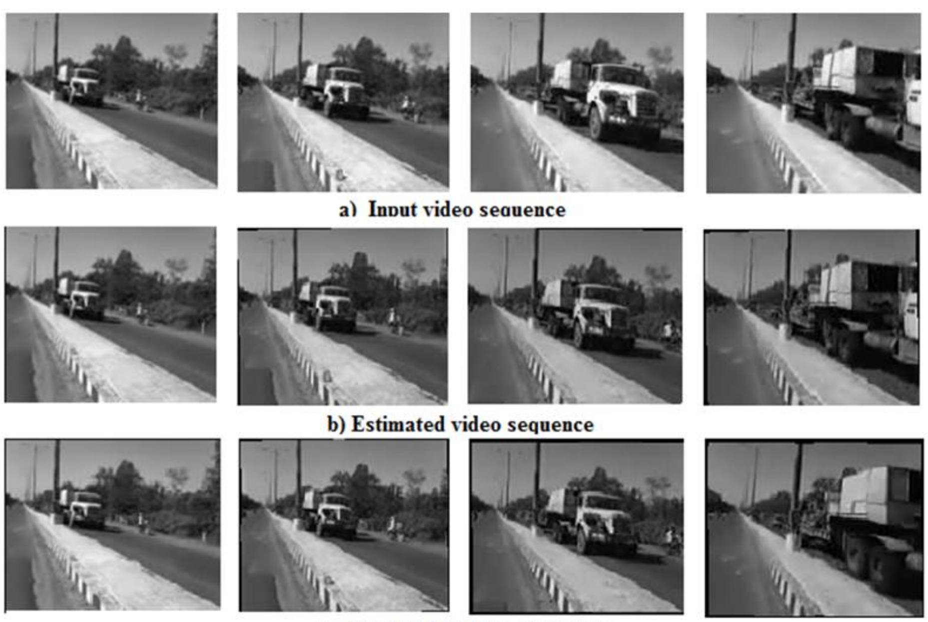


a) Input video sequence

b) Estimated video sequence

a) Stabilized video sequence

Figure 3. Performance of Proposed Algorithm for every 5th Frame for Corridor Video



a) Input video sequence

b) Estimated video sequence

c) Smoothed video sequence

Figure 4. Performance of Proposed Algorithm for Every 5th Frame for Highway Video

Table 1. Comparison of MSE for Input Video and Motion Estimated Video

Video	Mean Square Error		f1,f2	f2,f3	f3,f4	f4,f5	f5,f6	f6,f7	f7,f8	f8,f9	f9,f10	Peak to Peak diff.
HIGHWAY	Original video Before Stabilization		23.88	29.25	31.39	29.10	29.77	34.05	41.13	48.52	53.0	29.12
	After Stabilization	With simple mean filter	16.98	21.79	23.95	22.12	23.72	27.99	35.31	42.96	47.69	31.01
		With simple median filter	15.85	20.52	22.67	21.0	22.83	27.06	34.54	42.26	47.11	31.26
		with spline interpolation	8.45	7.79	8.448	13.38	13.88	7.052	30.08	20.58	20.7	23.05
		with Proposed bicubic interpolation	20.02	16.67	18.33	23.93	25.88	18.58	32.53	23.98	24.5	17.19
CORRIDOR	Original video Before Stabilization		15.78	10.87	8.35	13.93	21.99	25.26	31.36	22.4	20.06	23.01
	After Stabilization	With simple mean filter	11.19	7.28	5.36	10.18	17.78	21.35	28.03	18.12	15.59	20.75
		With simple median filter	10.53	6.97	5.09	9.79	17.12	20.73	27.55	17.45	14.92	22.46
		with spline interpolation	7.16	7.10	7.26	7.64	4.68	9.29	11.38	14.87	5.59	9.28
		with Proposed bicubic interpolation	21.05	21.62	22.93	24.89	22.12	27.14	27.28	27.81	23.68	6.76

Table 2. Comparison of SNR for Input Video and Motion Estimated Video

Video	Signal to Noise Ratio		f1,f2	f2,f3	f3,f4	f4,f5	f5,f6	f6,f7	f7,f8	f8,f9	f_9,f_{10}	Peak to Peak Diff
HIGHWAY	Original video Before Stabilization		5.68	4.64	4.34	4.71	4.64	4.06	3.37	2.86	2.67	3.01
	After Stabilization	With simple mean filter	7.99	6.23	5.69	6.19	5.82	4.94	3.93	3.23	2.97	5.02
		With simple median filter	8.56	6.62	6.00	6.52	6.05	5.10	4.02	3.29	3.00	5.56
		with spline interpolation	16.07	17.42	16.13	10.25	9.95	19.59	4.61	6.77	6.86	14.98
		with bicubic interpolation	6.78	8.15	8.35	5.73	5.33	7.43	4.27	5.79	5.78	3.98
CORRIDOR	Original video Before Stabilization		9.09	13.22	17.13	10.3	6.56	5.58	4.60	6.62	7.44	8.62
	After Stabilization	With simple mean filter	12.82	19.75	26.7	14.08	8.11	6.81	5.14	8.18	9.57	21.56
		With simple median filter	13.61	20.62	28.07	14.64	8.42	7.01	5.23	8.49	10	15.39
		with spline interpolation	20.06	20.22	19.69	18.8	30.81	15.65	12.67	9.97	26.7	20.04
		with bicubic interpolation	6.81	6.65	6.24	5.76	6.52	5.36	5.18	5.33	6.30	1.63

From Table 1 and 2 it can be evaluated that with proposed algorithm using a Bicubic interpolation MSE and SNR are more stabilized, while variation in MSE are very large with simple mean, median filters and Spline interpolation. It is clearly seen form Table 1 and 2 that minimum peak to peak difference of MSE is 17.19 for Highway video and 6.76 for Corridor video and minimum peak to peak difference of SNR is 3.98 for Highway video, and 1.63 for Corridor video which are obtained with proposed method using bicubic interpolation.

a) b) c) d)

Figure 5.Comparison of Optical Flow Method a) Original Frame b) Estimated Frame c) Difference of Frames for Horn and Schunck Derivatives [17] and d) Difference of Frames for Proposed Method

a) Input Frame Used by Feng Liu. [9] available at www.cs.wisc.edu/graphics/Gallery/Warp

b) Result of 2D Stabilization by Feng Liu. [9] with trimmed border area

c) Result of 2D Stabilization with proposed method

Figure 6. Comparison of Proposed 2D Stabilization Algorithm with Feng Liu [9]

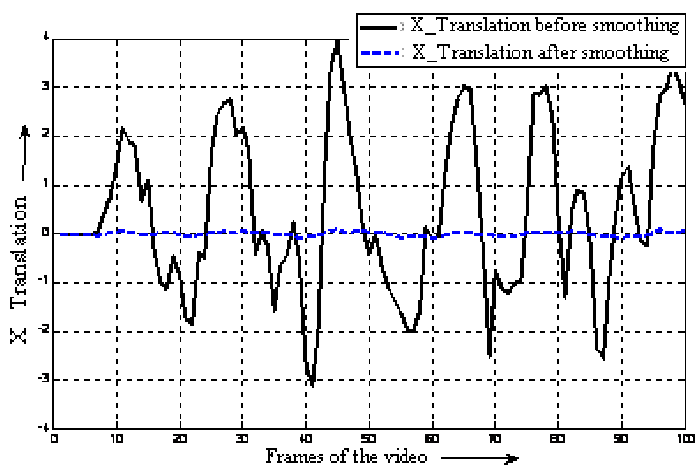


Figure 7. X Translation Before and After Motion Smoothening for Corridor Video

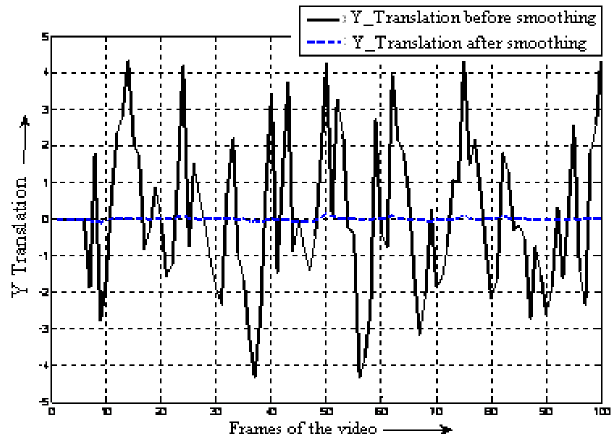


Figure 8. Y Translation Before and After Motion Smoothening for Corridor Video

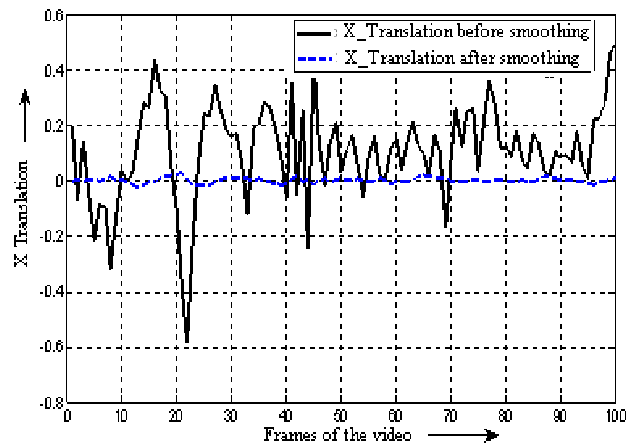


Figure 9. X Translation Before and After Motion Smoothning for Highway Video

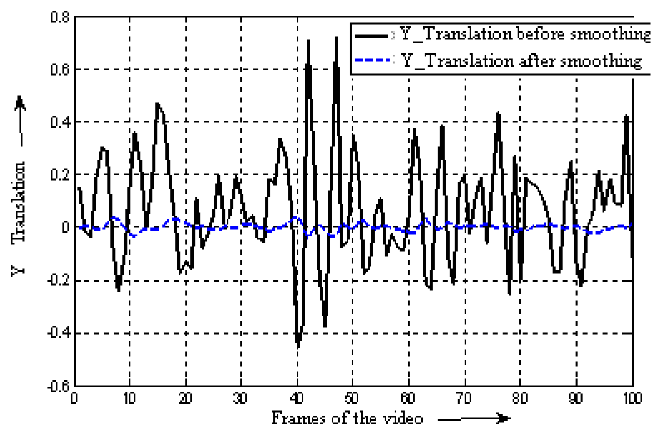


Figure 10. Y Translation Before and After Motion Smoothning for Highway Video

The accumulation error due to motion estimation has been minimized by using Gaussian kernel filtering for motion smoothening, to stabilize undesired translation in X and Y direction. Figure 7, 8, 9 and 10 shows stabilized translation in X and Y directions for 100 frames, before and after motion smoothening for Corridor and Highway video sequences. It can be seen that there is significant reduction in the undesired X and Y translations with the proposed method. Finally the rotation effects are removed by inverse filtering the smoothned affine parameters. Final stabilized results for the every 5^{th} frame of coorridor and highway videos are presented in Figure 3(c) and 4(c).

Proposed method is also tested on the videos used by Feng Liu [9], as shown in Figure 6. These videos are available in public domain of www.cs.wisc.edu/graphics/Gallery/Warp and are used for evaluation and comparison of proposed algorithm with latest existing work. As seen in Figure 6(b) that results of 2D stabilization by Liu. [9] method are slightly trimed at the boundaries of the frame, but proposed method generates full frame results with very small missing areas as shown in Figure 6(c).

5. Conclusion

In this paper a video stabilization algorithm for hand held camera videos is proposed. The results obtained with the proposed algorithm shows the stabilized motion in X and Y direction after motion estimation and compensation. The use of Taylor series improves the convergence rate and increases the efficiency of the motion estimation. To calculate the temporal derivatives 1-D separable filters are used which reduces the computation cost. The inter frame error between original input frames are compared with, inter frame error after motion estimation with mean filtering, median filtering, bicubic interpolation and spline interpolation. The method gives best stabilization with bicubic interpolation. It is found that peak to peak variation in MSE is reduced from 30 to 12 for Highway video and 23 to 7 for Corridor video with sequence of 10 successive frames. After motion estimation Gaussian kernel filtering is used for motion smoothening and finally the rotation effects are eliminated using the smoothed affine parameters and inverse rotation filtering. Method is capable of reducing the missing image areas significantly. There are few missing areas in the results as shown in Fig 3 and 4. In future these missing areas can be filled up to generate the full frame stabilized videos using video completion algorithm.

Acknowledgements

Authors especially want to acknowledge Mr. Feng Liu, and expresses thanks to made input videos globally available, and also grateful to D. P. Biswari for giving his valuable guidance time to time. Author also expresses sincere thanks to Dr.Ramakant Bharadwaj Prof. Truba Bhopal, for his expertise to make us learn the mathematics involved with research. Authors are grateful to everyone who have supported for our research work.

References

[1] Y. Matsushita, E. Ofek, Weina Ge, X. Tang and H. Y. Shum, "Full frame video stabilization with motion inpainting", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, **(2006)** July, pp. 1163-1178.

[2] H. C. Chang, S. H. Lai and K. R. Lu, "A robust and efficient video stabilization algorithm", ICME 04, International Conference on Multimedia and Expo, vol. 1, **(2004)** June, pp. 29-32.

[3] R. Hu, R. Shi, I. Shen and W. Chen, "Video Stabilization Using Scale Invariant Features", 11th International Conference on Information Visualization (IV'07) IEEE, **(2007)**.

[4] J. Yang, D. Schonfeld and M. Mohamed, "Robust Video Stabilization based on particle filter racking of projected camera motion", IEEE Transaction on, Circuits and Systems for Video Technology, vol. 19, no. 7, **(2009)** July, pp. 945-954.

[5] D. Pang, H. Chen and S. Halawa, "Efficient Video Stabilization with Dual-Tree Complex Wavelet Transform", EE368 Project Report, Spring, **(2010)**.

[6] R. Szeliski, "Image Alignment and Stitching: A Tutorial", Technical Report MSR-TR, 2004-92, Microsoft Corp., **(2004)**.

[7] H. Farid and J. B. Woodward, "Video stabilization and Enhancement", TR 2007-605, Dartmouth College, Computer Science, **(1997)**.

[8] O. Adda, N. Cottineau and M. Kadoura, "A Tool for Global Motion Estimation and Compensation for Video Processing", LEC/COEN 490, Concordia University, **(2003)** May 5.

[9] F. Liu, M. Gleicher, H. Jin and A. Agarwala, "Content Preserving Warps for 3D Video Stabilization", Int. Conf. Proc. ACM SIGGRAPH 2009 papers, New York, NY, USA: ACM, **(2009)**, pp. 1-9.

[10] C. Buehler, M. Bosse and L. Mcmillian, "Non-metric image based rendering for video stabilization", Proc. Computer Vision and Pattern Recognition, vol. 2, **(2001)**, pp. 609-614.

[11] A. Litvin, J. Konrad and W. Karl, "Probabilistic video stabilization using Kalman filtering and mosaicking", Proc. of IS&T/SPIE Symposium on Electronic Imaging, Image and Video Communications, vol. 1, **(2003)**, pp. 663–674.

[12] J. S. Jin, Z. Zhu and G. Xu, "Digital video sequence stabilization based on 2.5d motion estimation and inertial motion filtering", Real- Time Imaging, vol. 7, no. 4, **(2001)** August, pp. 357-365.

[13] G. Takacs, V. Chandrasekhar, D. Chen, S. Tsai, R. Grzeszczuk and B. Girod, "Unified real time tracking and recognition with rotation invariant fast features", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Fransisco, vol. 1, **(2010)** June, pp. 217-222.

[14] M. Pilu, "Video Stabilization as a Variational Problem and Numerical Solution with the Viterbi Method", Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, **(2004)**, pp. 625-630.

[15] J. R. Bergen, P. Anandan, K. J. Hanna and R. Hingorani, "Hierarchical Model based Motion Estimation", Proc. of Second European Conf. on Computer Vision, **(1992)**, pp. 237-252.

[16] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion", Int. Journal of Computer Vision, vol. 2, no. 3, **(1989)**, pp. 283-310.

[17] B. K. P. Horn and B. G. Schunck, "Determining optical flow", Artificial Intelligence, vol. 17, **(1981)**, pp. 185-203.

[18] Lucas Kanade.

[22] I. Multimedia, "Use Image Stabilization Gyroscopic Stabilizer", [online], URL http://www. websiteoptimization.com/speed/tweak/stabilizer.

[23] C. Morimoto and R. Chellappa, "Fast electronic digital image stabilization", Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, **(1996)** August 25-29, pp. 284-288.

[24] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "Performance of optical flow techniques", International Journal of Computer Vision, vol. 12, **(1994)**, pp. 43-77.

Authors

Prof. Paresh Rawat

Designation: Assistant Professor

B.E. 2000, M Tech, Ph. D* (Pursuing)

Dr. Jyoti Singhai

Designation: Associate Professor

Qualification: B.E., M. Tech., Ph. D

Research areas: Video & Image Processing