ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 8, August 2015

Robust Feature-based Digital Video Stabilization

Akshata Salunkhe, Sonal Jagtap

***Abstract*— A vigorous digital video stabilization approach is presented that provides both efficiency and robustness. In this approach, features of each frame are first detected by the oriented features from speeded up robust features (SURF) method, and then the corresponding features between consecutive frames are matched to speed up the motion estimation without any hardware. In addition, an improved motion smoothing method is proposed to smooth affine model based motion parameters by using filtering. This proposed method uses shaky input frames and stabilized output frames instead of original input frames to estimate motion parameters directly, allowing for more desirable motion parameters. Experiments with a variety of videos demonstrate that the proposed approach is both efficient and robust.**

***Index Terms*— Digital video stabilization, global motion estimation by SURF, motion smoothing, video warping.**

I. Introduction

The persons using camera are untrained, hence videos taken from hand held cameras suffers from undesirable motions due to unintentional camera shake during the scene capture time. These effects have weakened their performance significantly. The goal of video stabilization is to create a new video sequence where the motion between frames (or parts of a frame) has effectively been removed, and to synthesis a new image sequence as seen from a new stabilized camera trajectory. In the past decades, numerous researches have been done in the video stabilization field. There are two kinds of methods proposed to solve this problem: hardware approach and image processing approaches. Hardware approach, or optical stabilization, activates an optical system to adjust camera motion sensors. It is not broadly used due to their high cost and these are unable to process different kind of motions simultaneously. Another method used in stabilization is the image post-processing technique, which is our concern in this paper.

There are typically three major stages constituting a video stabilization process are, camera motion estimation, motion smoothing or motion compensation, and image warping. The video stabilization algorithms can be distinguished by the methods adopted in these stages. According to the computational complexity, most existing video stabilization algorithms fall into two types: offline post-process and real-time process, providing either high quality or robustness and efficiency. Real-time video stabilization is a very useful and attractive technique for compact video recording devices, e.g., cell phones, digital video cameras but the performance is limited because the adopted translational model is too simplistic, and the rotations are not taken into account. Therefore, the quality of stabilized sequence is no match for that of offline post-process. In contrast, offline post-process is more robust, and can provide high quality stabilized sequence by using feature-based motion estimation methods.

In this paper a robust video stabilization technique for removing the shakiness in the motion of hand held camera videos is proposed. The proposed technique uses global motion estimation using SURF to guess motion between the consecutive frames based on 2D affine transformation motion model. Then, a low-pass filter is applied to obtain the smoothed motion parameters. Finally, image warping warps the current frame according to the smoothed motion parameters and generates the stabilized sequence improving the efficiency of the video stabilization.

The rest of the paper is organized as follows. In Section II, the related work on video stabilization is summarized. The proposed video stabilization algorithm is shown in Section III. The performance of the approach is demonstrated in Section IV. Finally, Section V concludes the paper.

II. Related Work

Many methods for video stabilization have been reported over the past few years. Buehler *et al.* [1] proposed a novel approach by applying Image-based rendering techniques to video stabilization. The camera motion was estimated by non-metric algorithm. Image-Based Rendering was then applied to reconstruct a stabilized video and to smoothed camera motion. This method only performs well with simple and slow camera motion. The main drawback of approach is the difficulty of fitting motion models to complex motions. Matsushita *et al.* [2] developed an improved method for reconstructing undefined regions called Motion Inpainting and it was a practical motion deblurring method. This method produced good results in most cases, but it was strongly relies on the result of global motion estimation.

Battiato *et al*. [3] presented a VS technique based on the extraction and tracking of scale-invariant feature transform (SIFT) features through video frames using a feature-based motion estimation algorithm that tracks SIFT features extracted from video frames and then evaluates their trajectory to estimate inter-frame motion. Bosco *et al.* [4] have explored a novel approach for estimating the global motion between frames by analyzing the motion vectors obtained using block based motion estimation with the help of a dynamic curve warping, which is then incorporated in a system for VS for hand-held devices, giving enough robust results showing better shaky motion stabilization of randomly

Manuscript received August, 2015

***Akshata Salunkhe**, Department of Electronics and Telecommunication Engineering, STES's Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India.*

***Sonal Jagtap**, Department of Electronics and Telecommunication Engineering, STES's Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India.*

captured videos. Yang *et al.* [5] proposed the use of particle filters as a powerful and flexible tool to accurately model nonlinear physical systems to estimate the global camera motion between successive frames, thus by using them to estimate the affine transformation model of the global camera motion from corresponding feature points.

Zhou *et al.* [6] presented framework to solve the high-zoom video stabilization and completion problem by using a static low-zoom wide-view-angle camera and a synchro high-zoom active camera which will efficiently improve the accuracy of alignment among high-zoom views, which can help extracting more available high-resolution information for the completing. Mohamadabadi *et al.* [7] introduced a novel Radon transform based technique for VS that can efficiently deal with rotational in addition to translational motion and is robust to internal moving objects, occlusion and additive noise. Oreifej *et al.* [8] proposed a novel three-term low-rank matrix decomposition approach in which the turbulence sequence is decomposed into three components: the background, the turbulence, and the object, by focusing only on an image deformation because of the inherent confusion between the motion of the object and the motion caused by the turbulence.

III. PROPOSED APPROACH

The main challenge for any video stabilization algorithm is to distinguish the unwanted camera motion caused by unintentional camera motion from object motion and intentional camera motion. In the specified approach at first input video frames undergo feature point extraction and their matching. Later an affine transformation is performed to get motion model with affine parameters. Finally, motion smoothing is executed followed by video warping to achieve desired stabilized output video. In this way, this method stabilizes the shaky motion of the input video and increases its visual quality. The whole process is shown in Fig.1.Each block of Video Stabilization is described in detail next.

A. *Global Motion Estimation*

The first step in digital video stabilization is to determine how the camera is moving, this is called motion estimation. Camera motion is estimated by viewing sequential frames in a video feed. Here, global motion can be derived by finding features in subsequent image frames and then matching them to determine their start and end coordinates. Scale invariant feature transform (SIFT) and speed-up robust feature (SURF) are the most promising detectors due to good performance and have now been used in many applications. The SURF was conceived to ensure high speed in three of the feature detection steps: detection, description and matching [9].

At the beginning of the proposed method, SURF features are firstly extracted from two consecutive frames, and then key-points are matched. After the key-point matching, the affine transformation model is adopted in the proposed algorithm due to its high accuracy and the low computational complexity. The corresponding key-points for a frame of shaky video using the mentioned feature algorithms are illustrated in Fig. 2. From these extracted key-points, feature matching is performed between the consecutive frames

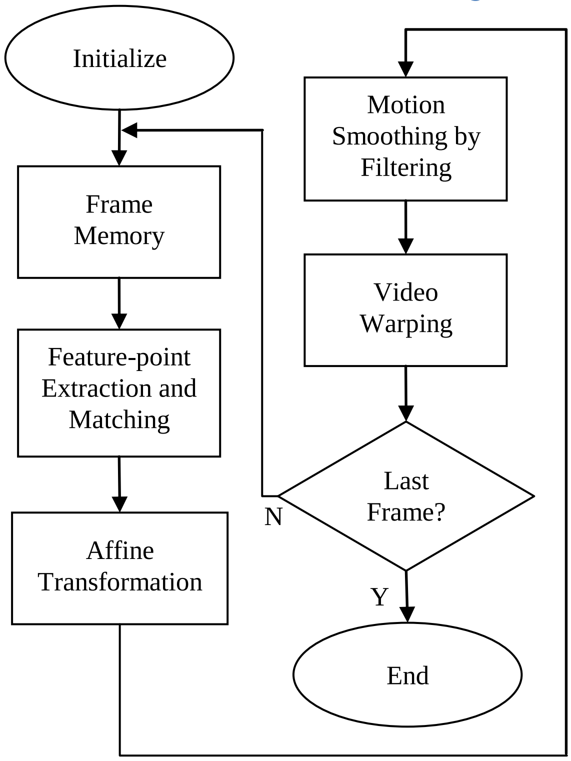

Fig. 1 The process of video stabilization algorithm

which are as shown in Fig.3.Given the matched key-points, the affine transformation parameters between the consecutive frames are estimated. It is any transformation that can be expressed in the form of a matrix multiplication (linear transformation) followed by a vector addition (translation). Consider pixel intensity values at point (x,y) in input image is to be transformed to (x',y') in output image. Now, the affine transformations are all transforms that can be written as

$$\begin{bmatrix} x' \\ \hline y' \end{bmatrix} = A \begin{bmatrix} x \\ \hline y \end{bmatrix} + B \qquad (1)$$

From above equation, the basic affine transforms can be calculated as follows:

- For pure translation transform, only matrix B is defined with matrix A=1.
- For pure scaling and rotational transform, only matrix A is defined but in this case matrix B=0.

An affine transformation does not necessarily preserve angles between lines or distances between points, though it does preserve ratios of distances between points lying on a straight line.

B. *Motion Smoothing by Filtering*

The goal of motion smoothing is to suppress high frequency jitters from the estimated global motion, and in the same time, obtain the intentional camera motion. Various smoothing methods have been used in video stabilization algorithms, such as Kalman filtering, Gaussian filtering, particle filtering, etc. Most of these motion smoothing methods always smooth out the accumulative global motion parameters with respect to a reference frame. As a result, the accumulative error is occurred and will be cascaded over time, which will lead to the failure of stabilization.

ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 8, August 2015

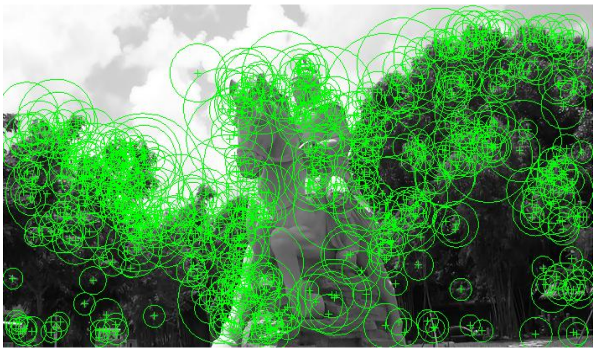

Fig. 2 Feature extraction using SURF algorithm

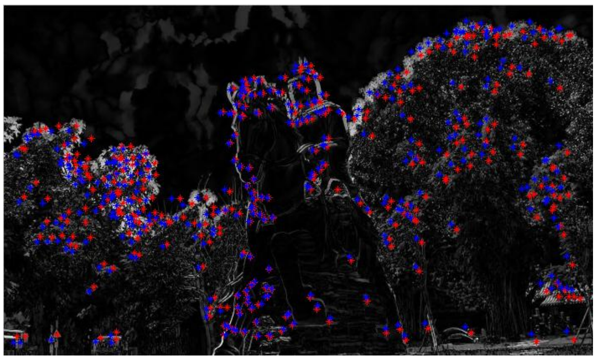

Fig. 3 Corresponding feature key-points matching between two consecutive frames (red and blue stars) using SURF algorithm

In motion smoothing for proposed video stabilization system, a moving average filter is employed to estimate the intentional motion by using both unstable input frames and stabilized output frames, which can further suppress high frequency jitters without increasing computational cost. The moving average filter is a simple Low Pass finite impulse response (FIR) filter commonly used for smoothing an array of sampled data/signal. It takes M samples of input at a time and takes the average of those M-samples and produces a single output point. It is a very simple LPF (Low Pass Filter) structure that comes handy for scientists and engineers to filter unwanted noisy component from the intended data. As the filter length increases (the parameter M) the smoothness of the output increases, whereas the sharp transitions in the data are made increasingly blunt. The moving average filter is given by

$$\overline{X}_k = \overline{X}_{k-1} + \frac{1}{n}\left[X_k - X_{k-n}\right] \qquad (2)$$

This is known as a moving average because the average at each k^{th} instant is based on the most recent set of n values. In other words, at any instant, a moving window of n values is used to calculate the average of the data sequence. Here, n=30 is taken.

The moving average filter is the simplest digital filter to understand. In spite of its simplicity, the moving average filter is optimal for a common task: reducing random noise while retaining a sharp step response. This makes it the most suitable type of filter for time domain encoded signals [10].

C. *Video Warping*

Video warping is the process of digitally manipulating a frame of video such that any shapes portrayed in the video have been significantly distorted. Warping may be used for correcting image distortion as well as for creative purposes (e.g., morphing). Thus, finally the motion smoothing frame can be warped from the original frame.

IV. EXPERIMENTAL RESULTS

The validation of the proposed algorithm is based on various real-life videos captured by a hand-held camera, as shown in Fig. 4.

Climbing: two men are up stepping, as well as many other objects come into video randomly through the camera lens which is a severe challenge for stabilization algorithm. It has 300 frames with resolution 396x704 and frame rate 30 fps.

Garden Visit: walking people in garden area and change in the scene containing variety of objects. It has high frequent jiggles and blur, and has 300 frames with resolution 396x704, frame rate 30 fps.

Monument: a statue which is standing at fixed position with free hand movements which have translational and rotational motions. It has 300 frames with resolution 396x704, frame rate 30 fps.

Volleyball Play: Some men are playing a volleyball changing scene while capturing ball hits by either sides of net. It has 300 frames with resolution 360x480, frame rate 30 fps.

Street: the changing scenes in a crossroad rapidly, accompanying by many moving objects, such as motorcycle, pedestrian. It has 300 frames with resolution 480x720 and frame 30 fps. Table I summarizes list of database video files including name of video file, resolution, data rate, size of file, frame rate and its duration.

Table I Video file details used for video stabilization

Input Sequence(.avi)	Resolution	Frame rate (Frames/sec)	Duration (sec.)
Climbing	396x704	30	10
GardenVisit	396x704	30	10
Monument	396x704	30	10
VolleyballPlay	360x480	30	10
Street	480x720	30	10

The proposed approach is tested on these five video sequences, which were taken by a digital camera in different scenes. These video sequences are with large translation, rotation and depth change. To evaluate the performance of the approach, the inter-frame transformation fidelity (ITF) measure is adopted which is used to measure the temporal smoothness:

$$ITF = \frac{1}{N_{frame} - 1} \sum_{K=1}^{N_{frame}-1} PSNR(K) \qquad (3)$$

where, N_{frame} represents the number of video frames. Higher the ITF value, more accurate is the estimation. As seen in above equation, ITF is a measure of peak signal to noise ratio (PSNR). The PSNR is computed just over the overlapping regions but for non-overlapping regions, a parameter inter-frame transformation fidelity (ITF) is used.

ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 8, August 2015

Fig.4 Sample videos taken for testing proposed video stabilization algorithm

The *PSNR* (*k*) is between two consecutive frames *(k, k+1)* which can be defined as:

$$PSNR(K) = 10\log_{10}\frac{I_{\max}^{2}}{MSE(K)} \quad (4)$$

where, $I_{\max}$ is the maximum pixel intensity and *MSE* (*k*) is the Mean Square Error between consecutive frames.

To check the effectiveness of the overall performance of the video stabilization system, a new parameter measure is presented. It is called as processing gain (PG). The PG is the measure of the ITF i.e. it is the ration of the difference between output ITF and input ITF to the output ITF. Thus, processing gain is given by

$$\%PG = \frac{ITF_{OUT} - ITF_{IN}}{ITF_{IN}} \times 100 \quad (5)$$

It is generally expressed in percentage. The quantitative evaluation of improved motion smoothing algorithm which uses original input shaky frames to smooth transformation parameters using ITF and processing gain using moving average filter, is given in Table II.

Table II Results of ITF and processing gain for proposed algorithm using moving average filter

Video Sequence	ITF Value in dB		% Gain
	Original Video	Stabilized Video	
Climbing	29.6749	30.2121	1.8102
GardenVisit	29.7886	30.1209	1.1155
Monument	30.8536	31.7276	2.8337
VolleyballPlay	29.7013	30.5439	2.8369
Street	32.9809	33.6842	2.1425

As can be seen from Table II, by smoothing the transformation parameters with output stabilized frames and input unstable frames, the improved smoothing algorithm can achieve higher ITF score of shaky video than that of stabilized video based on the same feature in each video sequence. It means that a more desirable and smoothed sequence can be obtained by the improved smoothing method giving better stabilization. It can be seen that ITF values for stabilized output by using moving average filter for given video database are validated compared to original video ITF values because greater is the ITF of stabilized video than original shaky video, better is the temporal smoothness of that video. For example, for input video sequence Climbing.avi, ITF value for original shaky video is 29.6749 and that for stabilized output video using moving average filter is 30.2121. For Street.avi, ITF value for original shaky video is 32.9809 and that for stabilized output video using moving average filter is 33.6842. Here, percentage ITF is also calculated to measure the processing gain of the system. Results of ITF and processing gain plots for the given 6 videos are shown in Fig .5 and 6.

Fig.5 shows the plot of the ITF values for stabilization effect using moving average filter. Since output ITF value is greater than input, result is validated. The plot of percentage processing gain is as shown in Fig.6. This figure shows the plot of percentage processing gain obtained from measure of ITF for moving average filter. Results are satisfactory and produced better ITF and % processing gain as compared to unstable input video.

V. CONCLUSIONS

The proposed video stabilization work performs better in looking for pixel correspondences in two neighbouring frames with motion stabilization. Spatial-temporal stabilization strategy of shaky motion strongly stabilizes video due to SURF descriptor and its feature matching efficiency. Motion stabilization effect is seen better by using moving average filter to smooth out the shaky motion in randomly captured video. The measurement of ITF and percentage processing gain achieve desired level of stabilization and guarantee the temporal smoothness of video. Thus, our system can handle videos with an extreme dynamic range of annoying motion improving the visual quality of amateur videos.

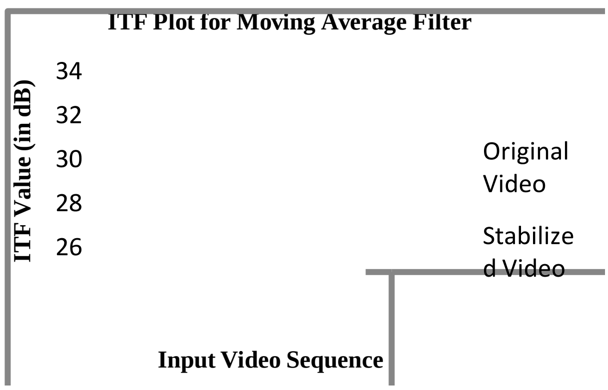

Fig. 5 Plot of ITF values of original videos and corresponding stabilized video

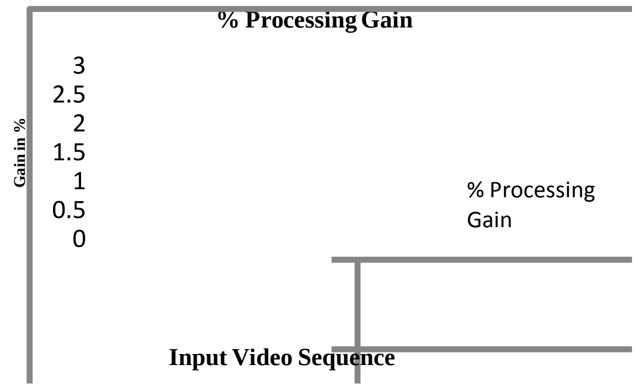

Fig. 6 Plot of processing gain values of original videos and corresponding stabilized video

If video frames are warped to achieve stabilization; our system may crop too much information if the given video is aggressively stabilized. This can be considered by solving the mentioned problems in our future work. One way to overcome these restrictions in the future can be to design stabilizing algorithms that are less sensitive to such changes and at the same time to achieve rough estimations of video stabilization with fewer assumptions, in less time.

ACKNOWLEDGMENT

I am indeed thankful to my guide Dr. S. K. Jagtap for her able guidance and assistance to complete this paper; otherwise it would not have been accomplished. I extend my special thanks to Head of Department of Electronics & Telecommunication, Dr. S. K. Shah who extended the preparatory steps of this paper-work. I am also thankful to the head & Principal of STES's, Smt. Kashibai Navale College of Engineering, Dr. A.V. Deshpande for his valued support and faith on me.

REFERENCES

[1] C. Buehler, M. Bosse And L. Mcmillan, "Non metric image-based rendering for video stabilization," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2001)*, vol.02, pp.609–614, 2001.

[2] Y. Matsushita, E. Ofek, W. Ge, X. Tang and H. Y. Shum, "Full-Frame Video Stabilization with Motion Inpainting," *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no.7, pp.1150-1163, Jul. 2006.

[3] S. Battiato, G. Gallo, G. Puglisi and S. Scellato, "SIFT Features Tracking for Video Stabilization," *IEEE 14th Int'l Conf. on Image Analysis and Processing (ICIAP)*, pp.825-830, Sep. 2007.

[4] A. Bosco, A. Bruna, S. Battiato, G. Bella and G. Puglisi, "Digital Video Stabilization through Curve Warping Techniques," *IEEE trans. on Consumer Electronics*, vol.54, no.2, pp.220-224, May 2008.

[5] J. Yang, D. Schonfeld and M. Mohamed, "Robust Video Stabilization Based on Particle Filter Tracking of Projected Camera Motion," *IEEE trans. on Circuits and Systems for Video Technology*, vol.19, no.7, pp.945-954, Jul. 2009.

[6] J. Zhou, H. Hu and D. Wan, "Video Stabilization and Completion Using Two Cameras," *IEEE trans. on Circuits and Systems for Video Technology*, vol.21, no.12, pp.1879-1889, Dec. 2011.

[7] B. B. Mohamadabadi, A. B. Khaligh and R. Hassanpour, "Digital Video Stabilization Using Radon Transform," *IEEE Int'l Conf. on Digital Image Computing Techniques and Applications (DICTA)*, pp.1-8, Dec. 2012.

[8] O. Oreifej, "Simultaneous Video Stabilization and Moving Object Detection in Turbulence", *IEEE Trans. on Pattern Analysis And Machine Intelligence*, vol. 35, no. 2, Feb. 2013.

[9] H. Bay, A. Ess, T. Tuyetelaars and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Elsevier Journal on Computer Vision and Image Understanding(CVIU)*, vol. 110, no. 3, pp. 346-359, Jun. 2008.

[10] Z. Huijuan and H. Qiong, "Fast Image Matching Based-on Improved SURF Algorithm," *IEEE Int'l Conf. on Electronics, Communications and Control (ICECC)*, pp.1460-1463, Sep. 2011.

Akshata Salunkhe was born in 1991. She has received B. E degree in Electronics Engineering from P.V.P.I.T., Budhgaon, Sangli and persuing her M.E degree in Electronics and Telecommunication with specialization in Signal Processing from Smt. Kashibai Navale College of Engineering, Vadgaon (BK), Pune in Savitribai Phule Pune University.

Prof. Sonal Jagtap was born in 1977. She received B.E. degree in Electronics and Telecommunication Engineering in 1999. She has also completed M.E. degree with specialization in Microwave Engineering from COEP, Pune University in 2002. She had completed her Ph.D. in Electronics from Shivaji University, Kolhapur in 2014. She joined as a lecturer in E&TC in Govt. Polytechnic Osmanabad in 1999 then to the Tuljabhavani College of Engg. in 2000. Presently she is working as a Associate Professor, Smt. Kashibai Navale College of Engineering, Vadgaon (BK), Pune since 2003. She is having 15 years of teaching experience. She has till now 2 national and 22 international journal publications in her credit. She has attended and presented 20 papers in International and 12 papers in National conferences.