A Robust Video Stabilization Method Based on Feature Point Extraction and Warping

Sonal.K.Jagtap, Himali.B.Ghorpade, Kanchan.L.Dombale

Abstract**: *In recent technologies there are various applications which include a camera joined to a moving platform, for example, cars, drones and Unmanned Aerial Vehicles (UAV). The moving platform may suffer from vibrations which may cause unwanted motion in recordings that can cause degradation of performance in various applications like surveillance, tracking and detection of object. Stabilization of video in various applications is an emerging research area nowadays. To remove the unwanted motion from video, the stabilization is necessary to preserve the important content present in the video. In this paper the feature points from recorded videos are detected and then these feature points are extracted and matched. The obtained feature points are smoothed by K means clustering, a mesh grid on every video frame is set up and every grid is warped by matching and comparing the features points, from original video frame with the smoothed and stabilized feature points. The reduced distortions in the video are estimated from various parameters. The efficiency of algorithm is compared in which the robust video stabilization algorithm based on feature extraction and mesh grid warping obtains better improvement in Inter-frame Transform Fidelity (ITF) factor than the traditional video stabilization algorithm.***

Keywords**: *Video stabilization, KLT tracker, Mesh grid warping***

I. INTRODUCTION

The vital difference between amateur-level video and professional video is the camera motion quality, hand-held videos are generally shaky and also expose spatial as well as temporal distortions and provide visual experience that is unpleasant whereas professional users take care of visual experience while recording and they use equipment such as dollies or steadicams to obtain proper stabilized video. These hardware are not practical for many situations, thus software based video stabilization is widely used; Also for driver safety applications a camera is mounted on the vehicle to catch the continuous recordings. Uneven surface of the streets and mechanical vibrations of the vehicle influence the nature of these recordings. The jitter makes unpleasant visual experience. Such movement of the camera additionally makes it hard to process and concentrate on imperative data from the pictures. Thus, the video should be rectified and stabilized. In this paper software based video stabilization is introduced which is efficient and robust and also provides good quality of results over a wide range of videos.

Revised Manuscript Received on October 15, 2019

Dr.Sonal.K.Jagtap, Professor and Head, Dept of E&TC Engineering in Smt Kashibai Navale College of Enginering, Pune-41.

Himali Babanrao Ghorpade, Assistant Professor in Smt Kashibai Navale college of Engineering, Pune-41

Kanchan Laxman Dombale, Department of Electronics and Telecommunication Engineering Smt Kashibai Navale college of Engineering, Pune-41

The video stabilization techniques can be divided into four different types like mechanical approach, optical approach, and digital approach. Digital stabilization approach is discussed in this paper. The ongoing digital video stabilization techniques are generally founded on stabilization of feature points and feature trajectory. In these techniques the distinct features are extracted and tracked accordingly. These features trajectories or feature points are then smoothed properly which follows the warping of video frames to supplant the first component features with the smoothed and balanced out positions as almost as could reasonably be expected.

II. RELATED WORK

Video stabilization procedures can be partitioned into two principle approaches: Global motion stabilization and feature based stabilization. In feature based stabilization techniques, with the best possible examination of a specific component, it permits to recover the global motion vector. The various features detectors used are Harris Corner Detector, Speeded Up Robust Features (SURF), Scale Invariant Feature Transform (SIFT). Feature based stabilization techniques extract feature points and form the trajectories by using feature points, and then smoothing of trajectories is performed. The motion in videos can be represented accurately by feature based stabilization techniques and also balance out recordings more fastly than the global motion method of stabilization.

In feature based techniques the analysis of optical flow is also performed. It comprises the arrangement of motion vectors that are determined between the two frames i.e. current and past frames which is generally used for depiction of video sequence [2, 3, 4].

The Steady flow motion model is analogous to the feature trajectory stabilization here motion is represented by the smallest pixel. In this method inspite of obtaining feature points and trajectories, it gathers a thick optical stream and the motion vectors are assembled and smoothed at a similar pixel areas. There is substantial computational complexity to compute a dense optical flow [3].

The Random Sample Consensus (RANSAC) [5] is utilized for adjustment of video. It requires excessively emphasis to acquire the outcomes, which backs off the entire handling and despite the fact that the expansions of RANSAC calculation are executed [6], they are very little noteworthy to fiercly build by and raise the overall execution.

The parameters of camera and 3D directions of an inadequate point are evaluated by using the Structure from motion Technique (SfM) technique. In this method to get a desired camera path they smooth the parameters and venture the 3D focuses on the balanced out camera. To manage the frame warping they used the 2D projected trajectories. Though they yield excellent outcomes on videos, SfM is sensitive to feature tracking failures [7].

The Global motion stabilization techniques also called as block based techniques, are generally used to calculate motion. The image is divided into small blocks and blocks of previous frame and current frame are compared with each other.

To examine the similarity between current and the previous frame blocks, the block matching criteria is utilized. To recuperate the movement vector for each block different coordinating techniques have been proposed: robust alignment based on Universal Image Quality Index (UIQI) [9] voting [10], and Mean Absolute Difference (MAD) [11]. They are not much suitable since the whole image should be fitted, in a block by block fashion which is very slow process

III. VIDEO STABILIZATION ARCHITECTURE AND METHODOLOGY

The video stabilization architecture shown in fig.1.consists of various blocks. Here video stabilization technique is implemented to decrease the distortions and undesirable camera motion. The video sequence used for stabilization is divided into various frames. These frames are then used for extraction of features points and its matching.

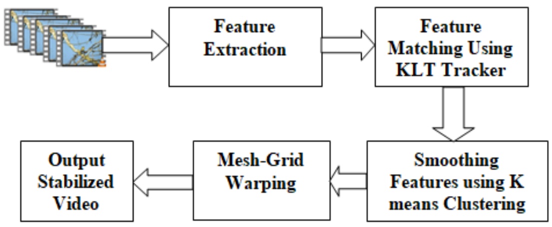

Fig.1. Architecture of video stabilization

Initially feature points are used for stabilization purpose, these feature points are detected, extracted and matched by KLT tracker through the frames of recorded video. Second the obtained feature points are smoothed, for this a k-means clustering technique is implemented. Then with these feature points, a mesh grid warping technique is applied to remove the distortions and stabilize each video frame. The obtained results show that this algorithm creates the video frames more stable than previous stabilization algorithms.

A. Feature Extraction

A very important step in stabilization is feature point extraction. The feature matching is best suited for successive pairs of frames. The proposed algorithm extracts the SURF (Speeded Up Robust Features) features. The integer estimation of the determinant of Hessian blob detector is utilized by SURF features, to find out the interest points. Its feature descriptor depends on the sum of the Haar wavelet response around the point of interest. These can also be calculated with the help of integral image.

The integral image is used to increase up the speed. It uses four pixels to compute the surface integral of any size from original image. Equation (1) represents the integral image

$$F(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(x,y) \quad (1)$$

The hessian matrix is used by the SURF to find the interest points. Given a point p=(x, y) in an image K, the Hessian matrix H (p, σ) at point p and scale σ, is represented as:

$$H(p,\sigma) = \begin{bmatrix} Kxx(p,\sigma) & Kxy(p,\sigma) \\ Kyx(p,\sigma) & Kyy(p,\sigma) \end{bmatrix} \quad (2)$$

The objective of a descriptor is to give a unique and powerful description of an image feature. Most descriptors are therefore processed in a nearby manner; hence a description is obtained for each and every point of interest determined previously.

Any mistake in feature extraction can degrade the stabilization performance. After extraction of features it is necessary to match and then smooth the features to achieve proper stabilization

B. Feature Matching using KLT Tracker

The Kanade-Lucas-Tomasi (KLT) tracker is used for matching the features points effectively across multiple frames. KLT utilizes spatial intensity data to coordinate the quest for the position that yields the best match. It is quicker than conventional strategies for examining at far less potential matches between the images. The KLT tracker will compute the tracked position in frame J for each feature point p in the frame I and the corresponding motion vector v. The intensity deviation should be zero between the feature points in frame I and its corresponding position in frame J. This KLT tracker is actualized for the most part to manage the issue that conventional image registration techniques are of high cost. This KLT feature tracker yields the best match. It is quicker than conventional methods.

C. Smoothing Features using K means Clustering

Smoothing the trajectories of features is an essential part in stabilization; a video sequence may consist of rolling shutter distortions. At first the feature positions extracted are distorted, as feature points are detected at various instants and at various rows in a frame. The K means clustering is used to smooth the feature points tracked by KLT point tracker. The K means clustering works as follows:

1) Let X be the set of data points and V be the set of centers.

$$X = \{x1, x2, x3, \dots, xn\} \quad (3)$$
$$V = \{v1, v2, v3 \dots, vn\} \quad (4)$$

2) Select randomly the cluster centers *'c'*.
3) Measure the length within each cluster center and data point.
4) Allot that data point to the center of cluster which is having minimum distance from all the cluster centers.
5) Again calculate the new cluster center by using the formula:

$$vi = \left(\frac{1}{ci}\right)\sum_{i=1}^{ci} xi \quad (5)$$

where, *'ci'* indicates the number of data points in i^{th} cluster.

6) Remeasure the distance between new obtained cluster centers and each data point.
7) Stop if no data point is reassigned, or else repeat from step 3.

Here K means clustering technique is used to cluster the feature points on object or background. Initially all feature points are divided into a K number, of different clusters and foreground features are determined. Then, K clusters are merged to obtain the two clusters: for object class and another for the background class. Lastly, the determination is done that if the object class consists feature points of a moving object and also the smoothing vectors of background feature points are used for the stabilization if the frame is bi-layered. Else all the smoothed vectors are used. Thereafter the image warping is very essential task performed which warps the image and achieves the stabilized video at the output.

D. Mesh Grid Warping

The input frame is applied with a mxn homogeneous mesh grid. The input frame is stabilized such that by observing the difference between the feature positions and its smoothed positions the corner points are warped accordingly. For each feature position p_i, the rolling-free smoothed position r_i is obtained as shown in fig.2.and then the feature positions are mapped with the smoothed positions as close as possible. To improve the stabilization performance effectively the feature reliability and each grid cell reliability is exploited. Stabilization is performed by mesh grid warping and at the output a stable balanced out video is obtained.

The new set of transformed corner points is denoted by V of all the grid cells. There is set of four grid cells as a 4 × 4 grid of cells is used then set of 25 such corner points are warped.

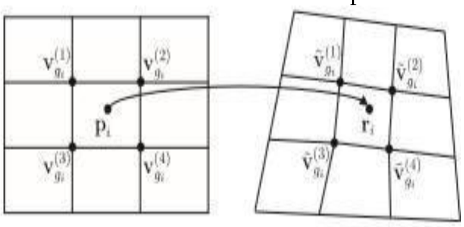

Fig. 2. Example of mesh grid warping [1]

The video stabilization algorithm implemented is explained (refer algorithm 1) in which detail procedure of algorithm is explained.

***Algorithm1**: The Video Stabilization algorithm*
1: Extract and Track the feature points.
2: **for each frame do**
3: **for each feature point Pi do**
4: Calculate its smoothed position ri
5: **end**
5: Cluster the feature points
6: Warp all the obtained smoothing vectors.
7: Output a stabilized frame
8: **End**

The feature points from the video frame are first extracted and then smoothed, where in smoothing each original feature point pi is smoothed with the feature point ri. The smoothing of feature point is then followed with the clustering of the feature points with K means clustering technique and finally to obtain a stabilized video frame, warping of all the obtained smoothing vectors is done. The experimental results will help to understand the effectiveness of the algorithm.

IV. RESULTS AND DISCUSSION

The video stabilization method is executed and simulated on MATLAB (R2013a), 64-bit (win64).The proposed method can be tested for different types of videos and videos with different scenarios with the '.avi' and '.mp4' extensions.

Video (a) Video (b) Video (c)

Fig.3: Video sequences used for stabilization

The simulation is performed on three different videos with various resolutions. Fig.3 shows the sample video sequences used for stabilization. Video (a) is shaky building and video (b) is shaky statue. Both the videos are recorded in camera whereas video (c) is shaky building which is recorded through cellphone. These videos having various resolutions and various parameters are shown in table.I

Table-I: Video with various resolutions used for stabilization

Input Video Sequence	**Resolution**	**Data Rate (kbps)**	**Size of Video (MB)**	**Frame Rate Frames/sec**	**Duration (sec)**
Video (a)	360x640	3000	1.53	30	10
Video (b)	396x704	3000	4.47	30	10
Video (c)	480x720	3094	2.28	30	5

The quantitative estimates like Root Mean Square Error (RMSE) and Peak Signal to Noise Ratio (PSNR), are calculated to determine the efficiency of proposed method. Results are calculated to get the answers of the following questions. How much increase in PSNR and reduction in RMSE is possible? How much is the efficiency of the proposed method? The comparison of the robust mesh grid

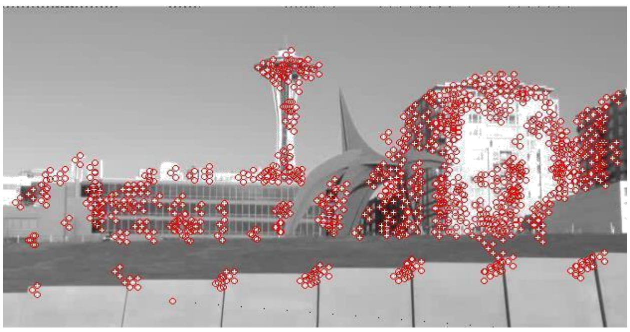

Fig.4 (b)Feature extraction and matching of video (a)

Retrieval Number: A9527109119/2019©BEIESP
DOI: 10.35940/ijeat.A9527.109119

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

warping method with kalman filtering method [12] is performed to determine the effectiveness of video stabilization method.

Fig.4 (b) Feature extraction and matching of video (c)

The feature extraction and matching of the shaky building and shaky statue can be observed in fig.4(a) and fig.4(b).The features indicated in red color and white color gives the feature point extraction and matching between two frames of the video where the white colored features are the feature extracted for frame 1 of same video and red colour are the feature extracted for frame 2 of same video.

The parameter Root Mean Square Error (RMSE) is computed to measure the minimum mean squared error. Lesser the mean square error greater is the stability of the video.

$$RMSE = \sqrt{\left[\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2\right]} \qquad (6)$$

The PSNR is Peak Signal to Noise Ratio, it is computed for the overlapping regions. The PSNR (k) is between two consecutive frames (k, k+1)

$$PSNR = 10\log_{10}\left[\frac{I^2max}{MSE}\right] \qquad (7)$$

The above obtained formulas for the calculations of performance parameters are implemented and the results are obtained as shown in table. II. The increase in the value of PSNR can be observed for original and stabilized video whereas the decrease in the value of RMSE for original and stabilized video is observed and analyzed

Table-II: Performance parameters of video sequences

Video Sequence	PSNR(dB)	RMSE
Original Video(a)	14.327	0.1922
Stabilized Video(a)	16.663	0.1499
Original Video(b)	10.592	0.3007
Stabilized Video(b)	11.672	0.2726
Original Video(c)	14.327	0.1929
Stabilized Video(c)	16.663	0.1498

The performance parameters viz. PSNR and RMSE for videos of both original video and stabilized video are calculated and it can be observed from table.II. that the stabilized video sequences gives better results as compared to original video sequences.

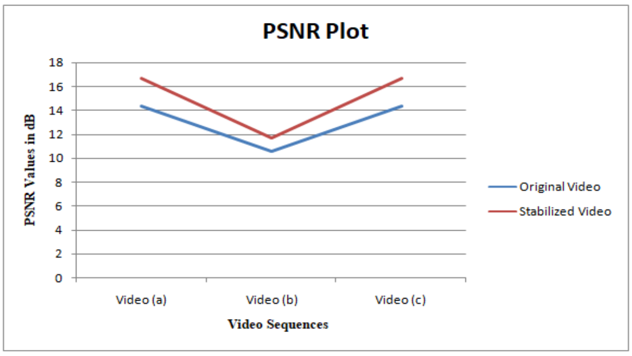

Fig.5 Graphical representation of PSNR values

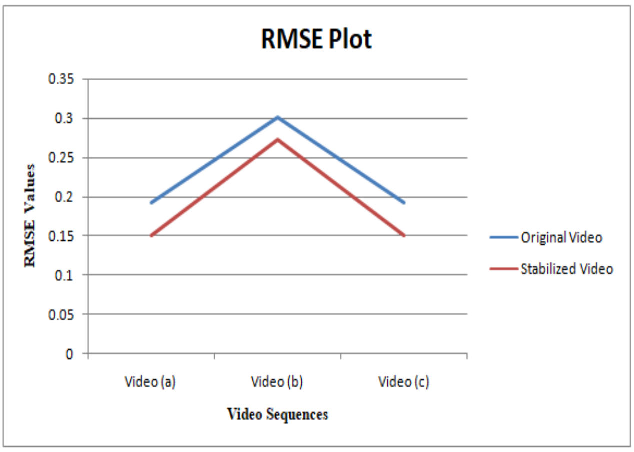

Fig.6 Graphical representation of RMSE values

Fig.5 and fig.6. represents the graphical representation of PSNR values and RMSE values of all the three videos. Both the original and stabilized videos are graphically compared with respect to PSNR and RMSE. It can be observed that PSNR of stabilized video is greater than the PSNR of original video also the RMSE of stabilized video is less that of original video.

The comparative analysis of the video stabilization algorithm with the kalman filtering method will give the efficiency of algorithm with respect to inter- frame transformation fidelity (ITF) parameter.

Table-III: Comparative analysis of Mesh Grid Warping Method with Kalman Filtering Method

Video	Methods used for stabilization	Original ITF(dB)	Stabilized ITF(dB)
Video(c)	Kalman Filtering Method[12]	25.65	27.54
	Mesh Grid Warping Method	27.15	30.31

Table.III represents the comparison of the proposed mesh grid warping video stabilization method with the Kalman Filtering Method [12]. The ITF of Stabilized shaky video through cell video having resolution (480x720) is compared with the kalman filtering method having the same resolution of the video. The performance of our approach in comparison with the ITF measure is more than kalman filtering, which means that the output video is more stable. This concludes that proposed robust feature extraction with mesh warping method gives better stabilization results as

Retrieval Number: A9527109119/2019©BEIESP
DOI: 10.35940/ijeat.A9527.109119

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

compared to other stabilization algorithms.

V. CONCLUSION

The feature extraction and mesh based warping video stabilization algorithm is based on reliable feature extraction which smoothes the features to obtain stabilization. The use of mesh grid warping achieves significant improvement to remove the shakiness in the arbitrarily captured video. The reduced distortions are estimated from parameters PSNR and RMSE and also the efficiency of algorithm is compared with the traditional algorithm. The robust video stabilization algorithm obtains almost 3dB of improvement in ITF factor and is robust to irritating motion, motion blurring and rolling shutter distortions enhancing the visual nature of amateur videos.

VI. FUTURE SCOPE

The future work will concentrate on developing the adaptive approach of stabilization and improve the matching accuracy as well as improve the efficiency of stabilization by improving the performance parameters used for stabilization i.e. PSNR, RMSE and ITF factor as well as to reduce the computational complexity.

REFERENCES

1 Y. J. Koh, C. Lee, C.S. Kim. "Video Stabilization Based on Feature Trajectory Augmentation and Selection and Robust Mesh Grid Warping". *IEEE Transactions on Image Processing*,2015

2 A. Goldstein and R. Fattal. "Video stabilization using epipolar geometry", *ACM Trans. Graph.*, vol. 31, no. 5, pp. 126:1–126:10, Aug 2012.

3 S.Liu, L.Yuan, P.Tan, and J.Sun,Steadyflow. "Spatially smooth optical flow for video stabilization", in *Proc. IEEE CVPR*, pp. 4209– 4216,Jun. 2014

4 Y.-S. Wang, F. Liu, P.-S. Hsu, and T.-Y. Lee. "Spatially and temporally optimized video stabilization", *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 8, pp. 1354–1361, Aug. 2013.

5 M. A. Fischler and R. C. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981

6 K Lebeda, J Matas, O Chum. "Fixing the locally optimized RANSAC",British*Machine Vision Conference, Guildford*,2012

7 F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3D video stabilization", *ACM Trans. Graph.*, vol. 28, no. 3, pp. 44:1– 44:9, Aug. 2009

8 V KovaazEvic, Z Pantic, A Beric, Jakovljevic R. "Block-matching correlation motion estimation for frame-rate up-conversion" Journal of Signal Processing Systems vol. 84 no. 2, pp. 283–292, 2016.

9 G Puglisi, S Battiato. "A robust image alignment algorithm for video stabilization purposes", *IEEE Trans Circuits System Video Technology* vol. 21, no. 10, pp. 1390– 1400,2011

10 D Bhujbal, BV Pawar. "Review of video stabilization techniques using block based motion vectors", *International Journal on Advance Research Science Engineering Technology* vol. 3 ,no. 3 ,2016.

11 A.U. Salunkhe, S.K. Jagtap, "Robust feature-based digital video stabilization", *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol 4, no.8 ,2015

12 A.Walha, A.Wali, A.M.Alimi. "Video Stabilization for Aerial Video Survelliance" in *AASRI Procedia Conference of Intelligent Systems and Control*,2013

13 B.D Lucas, T Kanade,. "An iterative image registration technique with an application to stereo vision", in: *Proceedings DARPA Image Understanding Workshop*, pp. 121–130 ,1981

AUTHORS PROFILE

Dr.Sonal.K.Jagtap received her Doctorate (Ph.D) degree in Electronics Engineering from Shivaji University, Kolhapur in 2014. She received M.E. degree with specialization in Microwave Engineering from COEP, Pune University in 2002 and B.E. degree in Electronics and Telecommunication Engineering in 1999. She is currently working as Professor and Head,Dept of E&TC Engineering in Smt Kashibai Navale College of Enginering,Pune-41.Her research interest include video and image processing, Artificial Intelligence and Machine Learning.

Himali Babanrao Ghorpade received M.E degree with specialization in Signal Processing in 2017 from Smt. Kashibai Navale College of Engineering, Pune. She received B.E degree in Electronics and Telecommunication Engineering from Sanjay Ghodawat Group of Institutes, Kolhapur in 2015. She is currently working as Assistant Professor in Smt Kashibai Navale college of Engineering,Pune-41 Her research interest include video and image processing.

Kanchan Laxman Dombale received M.E degree with specialization in Signal Processing in 2017 from Smt. Kashibai Navale College of Engineering, Pune. She received B.E degree in Electronics and Telecommunication Engineering from Dhole Patil College of Engineering, Pune in 2014. Her research interest include video and image processing.