International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-7S2, May 2019

Full frame Video Motion Detection and Stabilization using Mosaicing and Deblurring

N Sunny, M Srikanth, K Eswar

***Abstract*: Now-a-days, most of the videos captured are either from mobile phones or handheld video cameras. These videos are mostly shaky with undesired motion. The concept of video stabilization aims at removing this annoying shaky motion from the videos. In this process we estimate the motion of the camera, regenerate the motion of camera trajectory by removing the shaky component and complete the video with improvements like inpainting to fill the empty image areas. The pixel information of the nearby pixels is used to estimate the intensity of the missing pixels and filling the video frames. Though we can obtain a stabilized video, there are some parameters which can be improved such as key point detection, deblurring. So that we can get a superior quality video. Mosaicing is a parameter which is used to correct geometric deformations, video registration using video data and/or camera models and eliminating seams from video mosaics. The quality of the video after removing the shaky effect is further enhanced by using a deblurring algorithm. The smoothness of the video is improved by eliminating motion blur. It transfers and interpolates sharper image pixels from neighbouring frames to increase the sharpness of the frame. Using this process we create superior quality videos and reduce distracting vibrations and it is also used for improvising image quality in surveillance cameras**.

***Index Terms*: Deblurring, Frames, Image Stabilization, Motion blur, Mosaicing, Video Enhancement, Video Stabilization.**

I. INTRODUCTION

In the last years it has become a common thing to own a handheld video camera. The cameras have continued to become smaller and smaller in size. The videos captured these days are mostly from the handheld cameras or smartphones. The videos captured through smart phones and handheld devices are prone to camera motion and blur which in turn causing the video to have unwanted shakes and jitters. Hence stabilization of the video during post-production period can be done by using the digital approach [8] which makes use of the software. This software stabilization process is called as Digital video stabilization or Electronic image stabilization (EIS) [1]. This digital process of stabilization involves the steps of Estimating the trajectory of the camera motion, reducing the camera shake and getting the smoothened trajectory, using the process of inpainting to fill the empty spaces created due to the smoothing of trajectory. In the process of video stabilization, different parameters are used to stabilize the video and improve the video quality [6]. These parameters have a profound effect in the motion estimation stage which is the key to video stabilization. By choosing parameters such as deblur sensitivity the amount of blur that can be eliminated can be specified. These parameters can change from video to video because the condition in which the video is captured varies. Additional to this process we can use the image deblurring process to reduce the motion blur caused within the video by sharpening the pixels using the surrounding pixels. A video captured by using a handheld camera with a maximum length of 1 minute is taken as input for the application. The video captured can have background motion along with the motion of the camera such as a video captured while travelling in a car. The input video should be of a resolution 320 X 240 pixels and frame rate of 30fps. The output is a stabilized video with the same resolution similar to that of input video.

Revised Manuscript Received on May 28, 2019.

Nalluri Sunny, Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.
Mithinti Srikanth, Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.
Kodali Eswar, Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.

II. LITERATURE SURVEY

W. G. Aguilar etal.,[3] proposed an affine model to obtain a reliable transformation at an lower cost and deformation using the intentional motion of parameters and no feature points which solves the issues of phantom movements and minimizes the number of previous frames. His algorithm depends only on the last frame and is applied in the realtime without any decrease in performance. This mathematical transformation used is [3]

$$I_{s,p} = H_t . I_t \qquad (1)$$

where $I_{s,p}$ = [$x_{s,p}$, $y_{s,p}$,1]T and I_t = [x_t, y_t ,1]T are the coordinates of the interest points at the reference image and the uncompensated image, respectively, and H_t is the geometric transformation matrix.

W. Jiang .etal [5] proposed an algorithm to synchronize different videos into a single panaromic video. This algorithm comprises of two parts which are video frame alignment and spatial temporal seam finding. In the first part the set of videos are taken as input for which a common reference frame is selected and the remaining frames are aligned correspondingly. In this process the reference frame is selected based on the corresponding matching features. In the second part we combine the matching frames from different videos into an single panoramic video. This process is done by stitching the overlapping regions of the align frames to get the output video.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

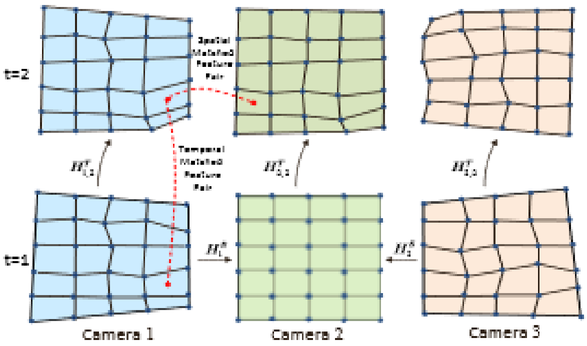

Figure 1 Group of videos stitched together [5]

L. Zhang, Q.-K. Xu etal [1] proposed an approach to find the set of optical image wraps for a given video sequence using shaky motion estimation and global formulation.
In the first method they extrapolate the shaky motion by using the future trajectories in the video sequences using the pyramidal lucas-kanade method [1]
$P = \{P_h\}_{h=1}^{M}$ Each trajectory P_h is composed of a set of nodes as the intersection between the trajectory and the corresponding frame.
In the second approach it uses the set of image wraps to remove the dissortions and the shakes in the stabilized sequence by the formula [1]

$$F(f_t) = F_m(f_t) + F_s(f_t) \quad (2)$$

Where F_m serves the steady motion of trajectories, and F_s controls visual distortion in image warping for the stabilized appearance.
J.kopf et. al., [2] proposed an approach for converting the frames into a less distorted cubical mapping representation and applying the inverse transformation for undoing the key frame rotations. The rotations are estimated with respect to the first frame by the formula [2]

$$R_{k_i} = \left(\prod_{j=1}^{i} R_{k_j}\right)^{-1} \quad (3)$$

For smoothing the trajectories in the inner frame rotations, the tracks from the visual trajectories which are plotted in the video are used. To reduce the residual jitter compensation, some flexibility is added to the motion model so that the video can be adapted to undo slight image deformations. After completing these steps extensions are added so that the performance is increased.
The extensions are:

- Reapplying Smoothed Rotations
- Two-pass Tracking
- Hyper lapse : It is calculated using the formula

$$v_j = median\{\cos^{-1}(P_j^i P_{j+1}^i) | f_i \le j < l_i\} \quad (4)$$

III. Methodology

The Figure 2 presents the flowchart for the proposed system which is used to remove the unwanted motion from the input videos.

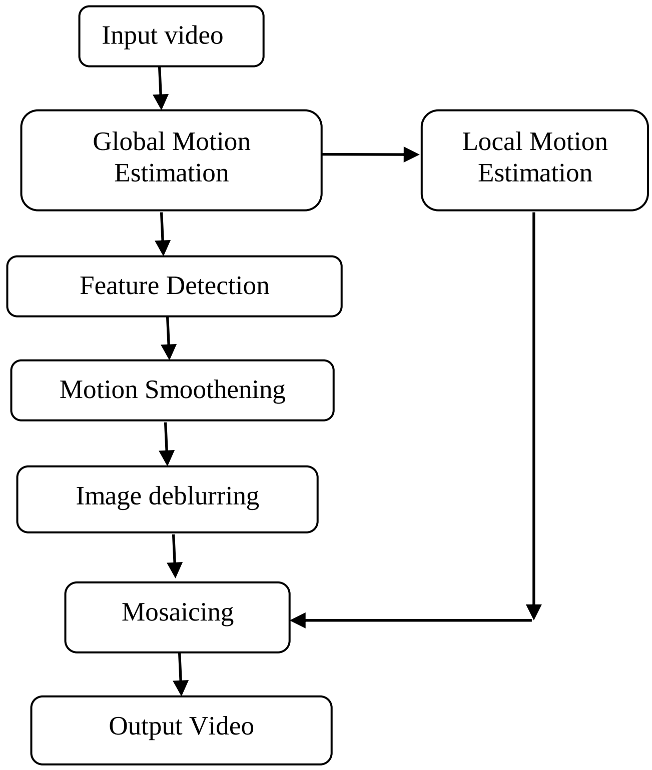

Figure 2 Flowchart of Proposed System

The input given to the system is a video containing unwanted jitter and shaky motions [12] which are caused due to the motion of the camera. The proposed method estimates the amount of foreground and background motion within the input video and reduces the unwanted motion present within the video to make it watchable. The proposed system is implemented in four modules which are motion estimation, motion smoothing, image deblurring and mosaicing. The motion estimation process is the initial step and is linked to the other steps directly. The output of one phase is directly given as input to the next phase. The final output video is generated after the mosaicing step which is the final step in the process. The output video is free of any unwanted motions and shakes caused due to the motion of the camera and lack of a stable environment.

A. Motion Estimation:

The Module of motion estimation is the initial and the important phase of video stabilization. This phase is divided into two parts which are Global motion estimation and the Local motion estimation [10]. Global motion estimation is a technique that attempts to map one image onto another with a simple four-corner pin. This differs from local motion estimation, which attempts to find where each individual pixel in the image is in the other image. Global motion estimation [10] is much cheaper to compute than Local motion estimation, but gives you less information about the image. We use the motion builder classes to estimate

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

the motion of a video. Since we have two types of motions we make use of two motion builders which are Motion Estimator Ransac L2 Builder and Motion Estimator L1 Builder.

The class MotionEstimatorRansacL2Builder estimates global motion between two 2D point clouds and also minimizes the L2 error. The L2 norm basically deals with minimizing the sum of the square of the differences between the target value and the estimated values. The class Motion Estimator L1 Builder is also use to estimate 2D global motion by minimizing the L1 error which provides the least absolute deviations. Both of these classes inherit many functions from the Motion Estimator Base class.

Algorithm:

Input: A Video captured with a handheld camera and having a resolution of 320 x 240 pixels and frame rate of 30fps.

Procedure:

1. Divides into frames I_t where t represents timestamp.
2. Frames $I_t, I_{t+1}, \dots\dots\dots$ are compared with each other for deviations in angles and transformations to estimate the motion.
3. Motion builders are used to find out the outliers within all the frames to estimate the deviation.
4. Frames excluding the outliers are saved and passed on to the motion stabilizer.

Output: The divided frames along with information about the outliers are passed onto next step.

B. Motion Smoothing:

The output from the motion estimation process is taken as the input for this phase. The motion stabilization is the processes of correcting the detected motion [9] and reducing the jitter within the video. This process is done by using the motion stabilizers which are TwoPassStabilizer and OnePassStabilizer which require the previous motion estimators. We can select (isTwoPass = 1) a one pass or a two pass stabilizer for this process.

The Two pass stabilizers usually result in much better quality but the two pass stabilizers are computationally slower. In the case of the other option, which is one pass stabilizer the result is not very good but it is faster compared to the previous. The stabilizer is the main part which sets the other options such as the source video file input and motion estimator to work properly. It also needs to cast the stabilizer to simple frame source video to read stabilized frames.

Algorithm:

Input: The frames from the motion estimation step are taken as input.

Procedure:

1. Wobble and jitter are identified in the frames.
2. High amounts of wobble and jitter invokes two pass stabilizer, else the one pass stabilizer is used.
3. Smoothened trajectory for the frames is established by performing the transformation on the frames.
4. Frames $I_t, I_{t+1}, \dots\dots\dots$ are transformed into $I'_t, I'_{t+1}, \dots\dots\dots\dots\dots$ using the formula

$$\prod_{t=0}^{i} I_{t+1}^{t} * I_{t}^{'t+1} \qquad (5)$$

Output: These smoothed frames are obtained.

C. Image Deblurring

Image deblurring is the third step in the process and aims at reducing motion blur. The process of deblurring is used to remove blur caused by defocus or by the motion of the objects. The blur in a video or image can be caused due to many reasons such as movement during the image capture process, by the camera or, when long exposure times are used. In the process of deblurring [11] the information from the nearby pixels is used to reduce the blur in a certain areas. We evaluate the "relative blurriness" of the image by measuring how much of the high frequency component has been removed from the frame in comparison to the neighbouring frames. Once relative blurriness is determined, blurry frames are sharpened by transferring and interpolating corresponding pixels from sharper frames.

Algorithm:

Input: The smoothed frames are taken as input from motion smoothing.

Procedure:

1. Finds the relative blurriness in the frames $I_t, I_{t+1}, \dots\dots\dots$
2. For I'_t consider the neighbouring frames I_{t-1} & I_{t+1} to sharpen the pixels and remove the blurriness using the deblurring formula

$$\hat{I}_t(p_t) = \frac{I_t(p_t) + \sum_{t' \in N} w_t^{t'}(p_t) I_{t'}(T_t^{t'} p_t)}{1 + \sum_{t' \in N} w_t^{t'}(p_t)} \qquad (6)$$

Output: The frames are obtained with reduced blur compared to input video.

D. Mosaicing

Mosaicing is used to fill the empty areas of the video caused due to the stabilization of the video. This is the final step in the process. The neighbouring frames [7] are all used for estimating and using the process of mosaic to filling the empty regions of the frame. The estimated information from the nearby pixels is used to complete the frame's empty areas. The frames after being filled are all stitched together [5] to form the required output video.

Algorithm:

Input: The deblurred frames are taken as input from the deblurring step.

Procedure:

1. M_t be a set of missing pixels in frame I_t fill in M_t using neighbouring frames $\{I_{t-k} \dots\dots\dots . I_{t+k}\}$ where k is timestamp.
2. Compute local motion from I_t to I'_t. It is estimated by

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

the optical flow between frames. A pyramidal version of Lucas Kanade optical flow computation[13] is applied to obtain the optical flow field

3. Fill empty parts by $I_t^{'}$ with the above motion estimated.
4. Loop above two points for all the empty areas.
5. After filling the empty areas stitch together all the frames into a single video of the same resolution and duration of the input video.

Output: The resulting stabilized video is obtained.

The output video is obtained after the completion of the fourth step which is mosaicking. The output video obtained contains fewer amounts of shakes and the efficiency of the stabilization is calculated using the number of frames stabilized. The output video is a video with higher quality than the input and is watchable.

IV. RESULTS

The video "inter_iit.avi" is the input taken and the process of motion estimation and motion correction are performed on this video. This produces a stabilized video which is done by using a Gaussian filter. The result of the stabilization is stored to the output path as "stabilized.avi". The input video scene is a parade of student in the ground captured by a handheld camera. The output video has reduced amount of jitter and wobble. The shakes present within the input video which are due to the camera motion are reduced. In comparison of the input and output frames of the video it can be observed that most of the parts of the output video have been smoothened by the algorithm so as to look good for the viewers. The stabilized video does not have any empty regions as the mosaicing is already applied to the video during the time of the stabilization and hence we get the final output. Some of the input and output video frames are shown below as a comparison.

Figure 3 Comparison between input and output video frames.

The computational time is proportional to the number of frames N and is also roughly proportional to the smoothness parameter k. Our proposed method handles the motion caused better than some other stabilization algorithms and reduces the movement of the video.

The performance of our implementation is based upon the number of frames into which the video is being divided. The efficiency of the process can be determined by calculating the ratio of number of frames stabilized to the total number of frames in the input video. The total number of stabilized frames is obtained by subtracting the number of outlier frames from the total frames present in the input.

Number of frames stabilized = [(total frames in input) – (number outlier frames)] (7)

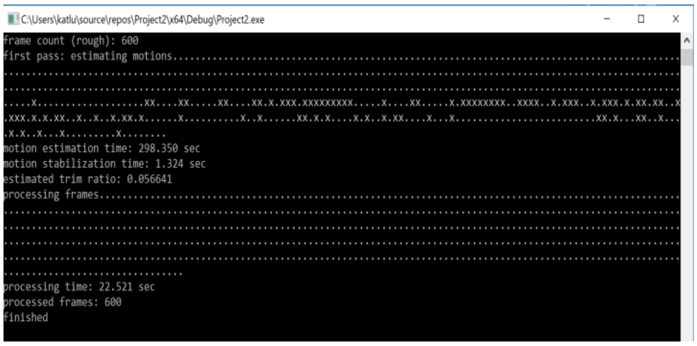

Figure 4 Computational times and process of proposed method

The efficiency of the stabilization process is expressed by using the number of frames stabilized. The number of frames stabilized divided by the total number of frames gives the efficiency of the process.

$$Stabilization\ efficiency = \frac{Number\ of\ frames\ stabilized}{Total\ number\ of\ frames} \qquad (8)$$

V. CONCLUSION

The videos containing jitter and shakes which make it hard for the viewers to watch it are taken as the input. The unwanted shakes and jitter from the video along with reducing the motion blur caused due to the foreground and background motions of the objects in the video are reduced. The proposed method also improves the smoothness of the video. The proposed method reduces the motion in four different steps. The feature points are used to estimate the motion present in between the frames. The frames are later transformed into a smooth trajectory by removing the outliers present in the video. The image deblurring is used to reduce the blur within the video or any blur which may be caused during the stabilization process. The final step of mosaicing stitches the frames together to form the output video. These resulting videos look natural and viewable.

In the future, the efficiency of this video stabilization algorithm can be increased by varying the values of the parameters such as nkps (number of key points to find in each frame), mosaicing, and LP based motion estimation, deblur sensitivity and so on for the input videos.

REFERENCES

[1] L. Zhang, Q.-K. Xu, H. Huang, "A global approach to fast video stabilization", IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 2, pp. 225-235, Feb. 2017.
[2] J. Kopf, "360◦ video stabilization," ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia), vol. 35, no. 6, 2016.
[3] W. G. Aguilar, C. Angulo, "Real-time model-based video stabilization for micro aerial vehicles", Neural Process.Lett., 2015.
[4] J. Dong, H. Liu, "Video Stabilization for Strict Real-Time Applications", IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 716-724, 2017.
[5] W. Jiang, J. Gu, "Video stitching with spatial-temporal content-preserving warping", Proc. IEEE Conf. Computer Vision Pattern Recognition Workshops (CVPRW), pp. 42-48, Jun. 2015.
[6] Paresh Rawat, JyotiSinghai. "Review of Motion Estimation and Video Stabilization techniques for hand held mobile video". Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011.
[7] A. Karpenko, D. Jacobs, J. Baek, and M. Levoy."Digital video stabilization and rolling shutter correction using gyroscopes". Stanford University Computer Science, Tech. Rep. CSTR 2011-03, September 2011.
[8] Brandon Smith, Li Zhang, HailinJin, AseemAgarwala. "Light Field Video Stabilization".IEEE International Conference on Computer Vision. 2009, pp. 341-348.
[9] Shuaicheng Liu, Yinting Wang, Lu Yuan, Jian Sun. "Video Stabilization with a Depth Camera". IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012, pp. 89-95.
[10] Mohammad JavadTanakian, Rezaei Mehdi, F. Mohanna. "Digital Video Stabilization System by Adaptive Fuzzy Kalman Filtering".Journal of Information Systems and Telecommunication. 2013, pp. 223-232.
[11] Xiaogang Chen, Feng Li, Jie Yang, and Jingyi Yu. "A theoretical analysis of camera response functions in image deblurring".University of Delaware, Newark, DE, USA, 2012.
[12] Matsushita, Y., Ofek, E., Weina Ge, Xiaoou Tang and Heung-Yeung Shum (2006). Full-frame video stabilization with motion inpainting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(7), pp.1150-1163.

AUTHORS PROFILE

Nalluri Sunny, working as an assistant professor in VR Siddhartha Engineering College, Vijayawada. completed M.Tech-CSE & B.Tech-CSE from JNTUK, Kakinada. Area of interest is Image Processing and Computer Networks.

Mithinti Srikanth, completed B.Tech- CSE form Acharya Nagarjuna University, Guntur and M.Tech-CSE from JNTUK, Kakinada. Area of Interest is Algorithms and Image Processing.

Kodali Eswar, working as an assistant professor in VR Siddhartha Engineering College, Vijayawada. completed M.Tech – Computer Science and Technology from AU College of Engineering, Visakhapatnam & B.Tech – CSE from Acharya Nagarjuna University. Area of interest is Network Security and Image Processing.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication