Video Stabilization Using Feature Point Matching

To cite this article: Shamsundar Kulkarni *et al* 2017 *J. Phys.: Conf. Ser.* **787** 012017

Video Stabilization Using Feature Point Matching

Shamsundar Kulkarni[1], D.S.Bormane[2] and S.L.Nalbalwar[3]

[1] Research Scholar, JNTU Kakinada, India.
[2] Principal,JSPM's Rajarshi Shahu College of Engg. ,Savitribai Phule Pune University , Pune,
[3] Professor & Head (E&Tc), Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, India.

Abstract. Video capturing by non-professionals will lead to unanticipated effects. Such as image distortion, image blurring etc. Hence, many researchers study such drawbacks to enhance the quality of videos. In this paper an algorithm is proposed to stabilize jittery videos. A stable output video will be attained without the effect of jitter which is caused due to shaking of handheld camera during video recording. Firstly, salient points from each frame from the input video is identified and processed followed by optimizing and stabilize the video. Optimization includes the quality of the video stabilization. This method has shown good result in terms of stabilization and it discarded distortion from the output videos recorded in different circumstances.

1. Introduction

Recently, the market of handheld camera has grown rapidly. However, video capturing by non-professional user normally will lead to unanticipated effects, such as image distortion, image blurring etc. Hence, many researchers focused on such drawbacks to enhance the quality of casual videos. [1]

Digital video stabilization is the process of removing unwanted movement from a video stream. Generally the processes of stabilization goes through three phases namely 1) Motion estimation 2) Motion smoothing 3) Motion compensation. The purpose of first phase is to estimate the motion between frames. After that the parameters of estimated motion which are obtained from the first phase will be sent to motion smoothening, where it removes the high-frequency distortion and calculates the global transformation, which is very important to stabilize the current frame [2]. Next, warping will be done by image composition for the motion compensation. These three-step frameworks are the essential steps in most of the video stabilization algorithms.

2. Literature Survey

Initial step in the video stabilization is selection of feature points. These feature points are corners.Corner detectors were developed in the late 1970's. Labeeb Mohsin Abdullah et.al. [1] had proposed an algorithm to stabilize the image sequence by using Harris corner detection technique. Harris Corner Detection is one of the fastest algorithms to find corner values [3]. Aleksandr Shnayderman et.al. had proposed SVD (Signature Value Decomposition)based gray scale image quality measurement. Edward Rosten et.al proposed faster and better machine learning approach to corner detection. This algorithm is used for

IOP Conf. Series: Journal of Physics: Conf. Series **787** (2017) 012017 doi:10.1088/1742-6596/787/1/012017

identifying Interest points in an image [4]. Yue Wang et.al. had proposed Real-Time Video Stabilization for Unmanned Aerial Vehicles. Elmar Mair et.al.proposed technique of Adaptive and Generic Corner Detection Based on the Accelerated Segment Test. C. Harris and M.J. Ephens proposed combined corner and edge detection to cater for image regions containing texture and isolated features, which is based on the local auto-correlation function. Mohammed A. Alharbi, had proposed Fast Video Stabilization Algorithm. [5] In this affine motion model is utilized to determine the parameters of translation and rotation between images. The determined affine transformation is then exploited to compensate for the abrupt temporal discontinuities of input image sequences. [6]

3. Methodology

Features from accelerated segment test (FAST) is a corner detection method which could be used to extract feature points and later used to track and map objects in many computer vision tasks. FAST is an algorithm proposed originally by Rosten and Drummond for identifying Interest points in an image. It is fast than many other well-known feature extraction methods, such as Difference of Gaussians (DoG) used by SIFT, SUSAN and Harris. An interest point in an image is a pixel which has a well defined position and can be robustly detected [7]. Interest points have high local information content and they should be ideally repeatable between different images.

3.1 Algorithm &Implementation

The Flow chart of video stabilization algorithm is shown in Fig.1. The total process of video capturing and stabilization has been performed in several steps. These steps are elaborated with simulation results.

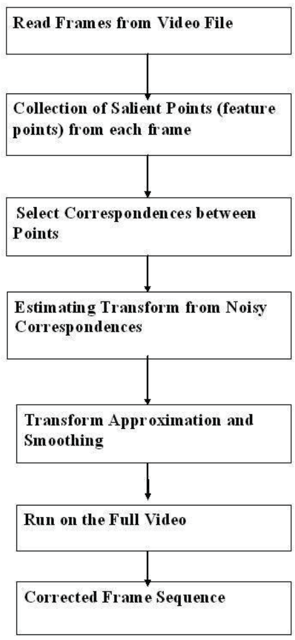

Figure.1 Flow chart of the of video stabilization algorithm

Step 1. Read Frames from Video File

In this step, first two frames of a video sequence are read. We read them as intensity images since color is not necessary for the stabilization algorithm and also using grayscale images improve the speed. Fig.2 shows both frames side by side. The data related intensity of images has been separated from color.

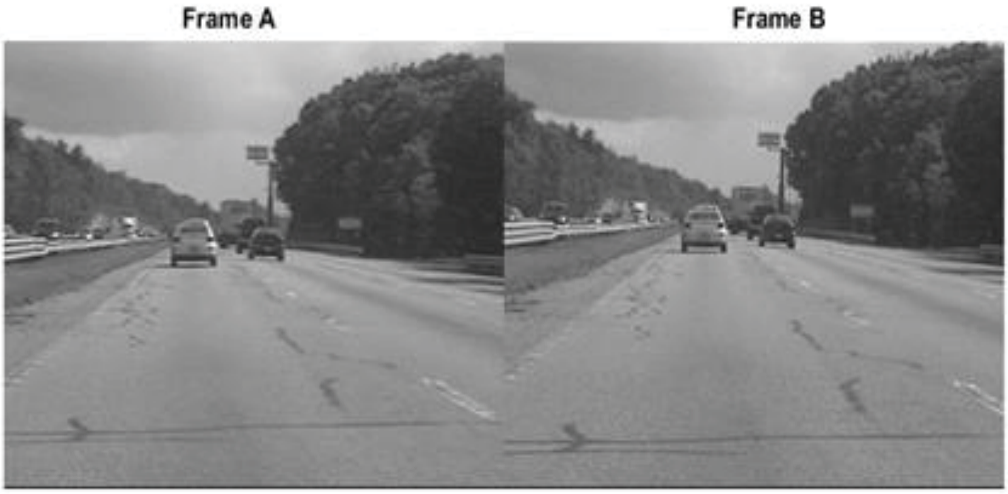

Figure 2. Reading the two frames

Step2 .Collection of Salient Points (feature points) from each frame

After reading two frames, Red-Cyan color composite is produced to illustrate the pixel wise difference between them as shown in figure 3. Our goal is to determine a transformation that will correct for the distortion between the two frames, for this Geometric Transform function can be used, which return an affine transform [8]. The input for this function shall comprise a set of point correspondences between the two frames. To generate these correspondences, collection of points of interest from both frames is performed [9]. For this purpose, FAST is corner detection algorithm is used. Fig.4 shows Image showing the interest point under test and the 16 pixels on the circle. It gives high measured quality video and less computational time. The detected points from both frames are shown in the fig.5 (a) &(b). Observe how many of them cover the same image features, such as points along the tree line, the corners of the large road sign, and the corners of the cars.

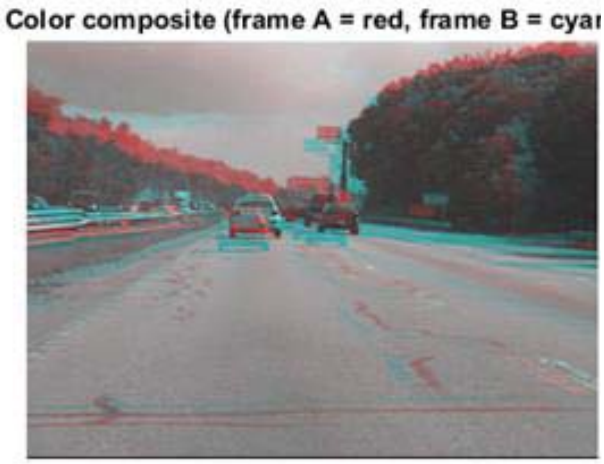

Figure 3. Red-Cyan color composite

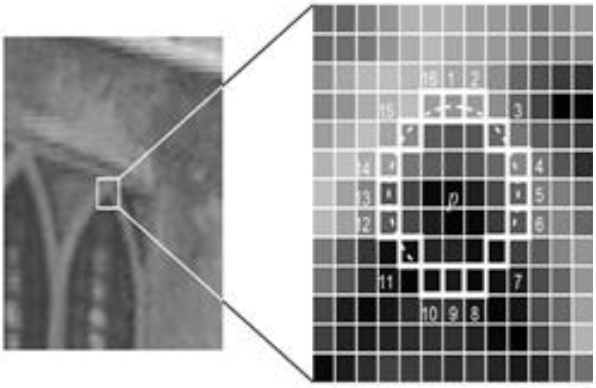

Figure 4. Image showing the interest point under test and the 16 pixels on the circle

IOP Conf. Series: Journal of Physics: Conf. Series **787** (2017) 012017 doi:10.1088/1742-6596/787/1/012017

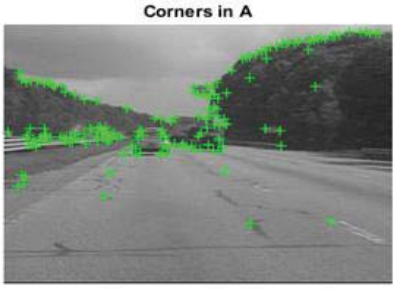

Figure 5. (a) corners in frame A

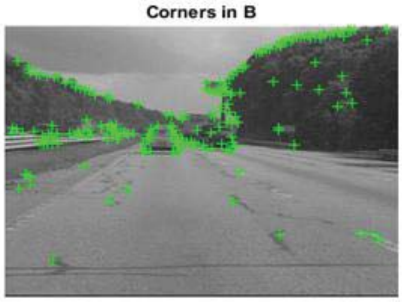

Figure 5. (b) Corners in frame B

Step 3. Select Correspondences between Points

In this step, correspondences between the points derived in the step 2 should be established. For each point, we extract a Fast Retina Key point (FREAK) descriptor centered around it [10]. The matching cost we use between points is the Hamming distance since FREAK descriptors are binary. Points in frame A and frame B are matched putatively. Note that there is no uniqueness constraint, so points from frame B can correspond to multiple points in frame A. Fig.6 shows the same color composite given above, in this points from frame A are shown with red colour, and the points from frame B in green colour. Yellow lines are drawn between points to show the correspondences. Many of these correspondences are correct, but still there is significant number of outliers.

Step 4. Estimating Transform from Noisy Correspondences

Many of the point correspondences obtained in the previous step are identified with limited accuracy. To rectify this problem, Random Sample Consensus (RANSAC) algorithm is used which is implemented in the Geometric Transform function. It is observed that the inliers correspondences in the image background are not aligned with foreground. The reason behind this is the background features are far enough those act as if they were on an infinitely distant plane. We can assume that background plane is static and will not change dramatically between the first and second frame, instead, this transform is capturing the motion of the camera. Thus correcting process will stabilize the video. Furthermore, as long as the motion of the camera between frame A and frame B is minimize or the time of sampling the video is high enough, this condition is maintained. The RANSAC algorithm is repeated multiple times and at each run the cost of the result is calculated by projecting frame B onto frame A via Sum of Absolute Differences between the two image frames. Fig.7 shows a color composite showing frame A overlaid with the re-projected frame B, along with the re-projected point correspondences. It is clear from this figure that the results are favorable, with the inliers correspondences nearly exactly coincident. The cores of the images are both well aligned, such that the red-cyan color composite becomes almost purely black-and-white in that region.

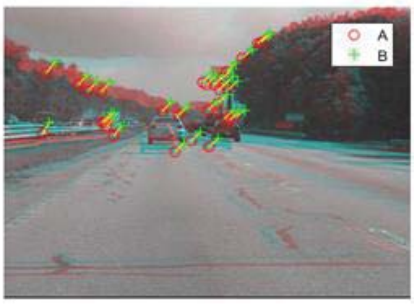

Figure 6. Correspondences between Points

Figure 7. color composite showing frame A overlaid with the re-projected frame B

IOP Conf. Series: Journal of Physics: Conf. Series **787** (2017) 012017 doi:10.1088/1742-6596/787/1/012017

Step 5. Transform Approximation and Smoothing

We could use all the six parameters of the affine transform, but, for numerical simplicity and stability, we choose to re-fit the matrix as a simple scale-rotation-translation transform. This has only four free parameters as compared to the full affine transform which are one scale factor, one angle, and two translations. Finally reconstructed S-R-t transform output is shown in Figure 8.

Step 6. Run on the Full Video

The above procedure of estimating the transform between two images has been performed in the MATLAB® function cvexEstStabilizationTform. The function cvexTformToSRT also converts a general affine transform into a scale-rotation-translation transform. The smoothed video frame is shown in fig.9 by using Video Player in Matlab.

Figure 8. Color Composite of Affine and s-R-t Transform output

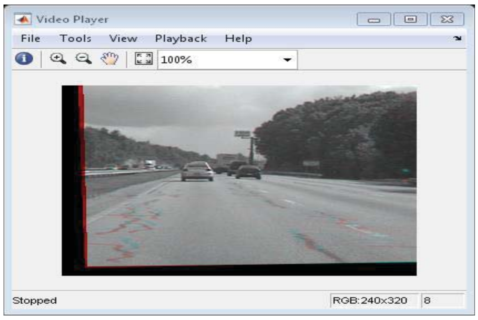

Figure 9. Camera movement to stabilize video

Step 7. Corrected Frame Sequence:

During computation, we computed the mean of the raw video frames and mean of the corrected frames. These mean values are shown side-by-side in fig.10. The left image shows the mean of the raw input frames, proving that there was a great deal of distortion in the original video. The mean of the corrected frames on the right, however, shows the image with almost no distortion. Foreground details have been blurred because of the car's forward motion. This shows the efficacy of the stabilization algorithm.

Figure.10 Mean of the raw video frames

4. Result and conclusion

In this experimentation two shaky videos are considered. FAST detection method is applied for feature point detection. Quality factor and computational time are analyzed. Video 1(shaky _car) has the following specifications: the length of 4 seconds, width x height of 320x240, frame rate of 30 frames/second and size of 1.2 MB. . Video 2(shaky _Highways) has the following specifications: the length of 5 seconds, width x height of 320x240, frame rate of 30 frames/second and size of 1.4 MB. This technique provides logical and computationally efficient approach in terms of stabilizing high jitter videos suffered from distortion. The FAST detection algorithm presented in the work provides a fast and robust stabilization system and improves real-time performance. The quality and computational time are measured and compared for sample videos as shown in Table 1.

Table 1. Performance parameters

Sr.No	Sample Video	Computational Time (s)	Quality Value
1	Shaky _video1	8.81	22.20 %
2	Shaky _video2	9.18	39.21%

5. Conclusion and Future scope

The use of FAST detection method shows improved result in terms of stabilization. FAST detection technique is useful in enhancing the quality of low-grade video surveillance cameras. FAST detection technique is particularly helpful in identifying people, license plates, etc. from low-quality video surveillance cameras. In future, we can find better feature detector to overcome the consequences of extreme shaking of handheld camera in real time implementation for video stabilization.

References

[1] Aleksandra Shnayderman, Alexander Gusev, and Ahmet M. Eskicioglu"An SVD-Based Grayscale Image Quality Measure for Local and GlobalAssessment ",IEEE 15(2), 2006.

[2]B.-Yu. Chen, K.-Yi. Lee, W.-T. Huang, J.-S. Lin, "Capturing Intention-based Full-Frame Video Stabilization," Computer Graphics Forum, Vol. 27, No. 7, p.1805 - p.1814, 2008.

[3] C. Harris and M.J. Stephens, "A combined corner and edge detector", Proc of Alvey Vision Conference, pp 147–152, 1988.

[4] Edward Rosten, Reid Porter and Tom Drummond, "FASTER and better: A machine learning approach to corner detection" in IEEE Trans. Pattern Analysis and Machine Intelligence, 2010, vol 32, pp. 105-119

[5] C. Morimato and R. Chelleppa, "Evaluation of Image Stabilization algorithms", Proc. of IEEE int. Conf. on Acoustics, speech and signal processing, Vol-5,pages 2789-2792, 1995.

[6]Jie Xu, Hua-wen Chang, Shuo Yang, and Minghui Wang "Fast Feature-Based Video Stabilization without Accumulative Global Motion Estimation" IEEE 2012.

[7]Ken-Yi Lee Yung-Yu Chuang Bing-Yu Chen Ming Ouhyoung "Video Stabilization using Robust Feature Trajectories" 2009 IEEE 12th International Conference on Computer Vision (ICCV).

[8]Keng-Yen Huang, Yi-Min Tsai, Chih-Chung Tsai, and Liang-Gee Chen "Feature-based Video Stabilization for Vehicular Applications" 2010 IEEE 14th International Symposium on Consumer Electronics.

[9] Labeeb Mohsin Abdullah, Nooritawati Md Tahir & Mustaffa Samad , "Video Stabilization based on Point Feature Matching Technique", 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012).

[10]Xie Zheng, Cui Shaohui, Wang Gang, Li Jinlun "Video Stabilization System Based on Speeded-up Robust Features" International Industrial Informatics