Comparative Analysis on Feature Descriptor Algorithms to Aid Video Stabilization and Smooth Boundary Reconstruction using In-painting and Interpolation

Mitali J. Patel
ME Student CSE Dept.
Parul Institute of Technology, Limda

Neha Parmar
Asst. Prof., CSE Dept.
Parul Institute of Technology, Limda

Maheshwari Nilesh A.
Student EC Dept.
Parul Institute of Technology, Limda

ABSTRACT

This paper presents a fast and efficient video stabilization method based on the Speeded-up robust features (SURF). We adopted speeded-up robust features as feature descriptor, which are extracted and tracked in each frame .This extracted features are matched through SURF, matched features were used to estimate the geometric transformation between the frames. Finally estimated transformation is applied to the frames to produce a new stabilized frame pair. After the geometric transformation is carried out, the resultant frames are almost stable. But the boundary region of stabilized frames requires a lot more attention as they are said to be black and some sort of filtering and inpainted work needs to be performed for better results and reconstruction of the obtained stabilized frames. Hence in the obtained stabilized frames, we need to estimate the exact location of the regions at the boundary where inpainting work is to be carried out. Experimental results illustrate superior performance of the SURF based video stabilization in terms of accuracy and speed as compared to other state of art algorithms based stabilization method.

Keywords
Video Stabilization, Inpainting, SURF.

1. INTRODUCTION

Recently we have saw that the market of hand-held video cameras have growth rapidly in popularity. But the videos retrieved from such devices are affected by unwanted camera shakes and jitters, resulting in video quality loss. Hence video stabilization is a technique that is used to improve the video quality by removing the unwanted camera movements due to hand shaking and unintentional camera shake. The aim of video stabilization is to smooth the blurred video that is caused by undesired movement in camera. In the past decades, numerous researches have been done in the video stabilization field. Generally, the process of video stabilization consists of three major steps: motion estimation, motion smoothing, and video completion.
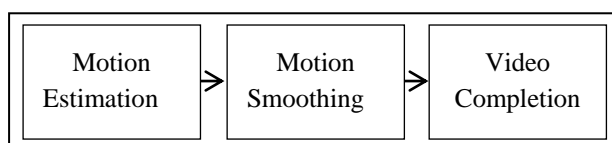

Fig 1: General Video Stabilization Process

The goal of motion estimation is to make an attempt to establish the correlation and the correspondence of the objects/contents between successive frames in a video, in such a way that camera motion in the video can be estimated and represented. To achieve this goal, two main-stream motion estimation categories are: optical flow estimation and feature matching. The motion between two consecutive frames is characterized by optical flow either on a pixel basis or on a frame basis. On the contrary, the latter method that extracts geometrically invariant key points from each frame, finds corresponding feature pairs, and then derives the transformation between two consecutive frames. In our investigation, on the one hand, integrating the optical flow estimation into the video stabilization process usually trades the efficiency for the elegant formulation to obtain and analytic solution. In Optical flows estimation method, to stabilize the camera motion in 3-D space during the process of mosaic construction. and In our proposed method to use of the global motion estimation and the local motion estimation together to remove undesired motions. On the other hand, the video stabilization method based on feature matching usually benefits from the efficiency. We, therefore, base the proposed video stabilization process on the feature-based (block-based) scheme.

In regard with motion smoothing, there also have been a considerable number of the existing algorithms. In [11], [12], Matsushita et al. proposed the new transformation matrix by averaging the set of transformation matrices associated with certain number of neighboring frames. In [14], Gleicher and Liu attempted to reconstruct the motion path of a virtual camera, which is used as a guide to improve the camera dynamics in a video. In [13], Chen et al. proposed to establish the trajectory for a target object and then interpolate a smoothen trajectory, which is used to produce the stabilized video sequence. Since the method described in [11], [12] is suitable for the integration of the feature matching step into the stabilization process. We, therefore, utilize the local displacement method for smoothing the camera motion.

In regard with video completion, there are two mainstream methods: (1) extraction of the valid sub-frame from the resultant video; (2) filling of the missing pixels. As the result that the latter method works well for full-frame resolution, we propose to fill in the missing pixels and image regions.

Un-stable Frame   Stable Frame

Fig 2: Video Stabilization

In general, video stabilization consists of three steps: motion estimation, motion smoothing, and image warping. Firstly, motion estimation step estimates the motion between consecutive frames based on a specific 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.

2. DIFFERENT METHODS

There are two types of stabilization: Hardware and software based stabilization.

In hardware-based stabilization sensors and lens are used to reduce the movement of cameras. However, these hardware-based systems fail to provide desired stabilization to compensate for complex camera motions and jerks. Therefore, to obtain stable videos, post- processing video stabilization is required. Video stabilization removes the undesired motion from input video by accordingly warping the images [7].

In software-based video stabilization video is stabilized in three steps as shown in fig. 1. Firstly motion estimation is done between two sequential frames i.e. the previous and current frames. Motion compensation provides the computation of global transformation to stabilize the frame content. Steps followed for motion compensation are:

- Split video into two parts i.e. static parts and moving parts.
- Get the data of the previous frame on the basis of motion vectors.
- Use filters to obtain the forecast difference block between previous and current frame.
- Static parts and the forecast difference block are combined and new image is regenerated.
- And finally warp the current image with the previous image on the basis of transformation [7].

3. FEATURE EXTRACTION METHODS

3.1 Harris Corner

Harris corner detector is detecting the corner values. It is based on the local auto- correlation function of a signal. The local auto correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. Also it is popular because it is invariant to translation, rotation and illumination change. This detector is most repetitive and most informative. The disadvantage of this detector is it is not invariant to large scale change[10].

3.2 FAST Feature

Features from accelerated segment test (FAST) is a corner detection method, which is many times faster than other existing corner detectors due to the high speed, it is a good choice for finding keypoints in real-time systems that match

visual features, for example, real-time video stabilization. Although FAST is superior in computation, it does not include an orientation operator which is involved in many keypoint detectors, such as SIFT and SURF. This defect leads to low robust to image noise. By adding an orientation operator, FAST can overcome this weakness effectively [9].

3.3 ORB

ORB (Oriented FAST and Rotated BRIEF) can be a computationally efficient replacement to SIFT. ORB has similar matching performance with that of SIFT, and is less affected by image noise. Moreover it is capable of being used for real-time performance. The main motivation of ORB is to enhance the efficiency of many common image matching applications, e.g., to enable low-power devices to perform panorama stitching and patch tracking without GPU acceleration, and to reduce the time for feature-based object detection on standard PCs. Arguably, ORB performs as well as SIFT and SURF on these tasks, while being almost two orders of magnitude faster. Therefore, the ORB feature is very suitable to be used for the global motion estimation of video stabilization.

3.4 MSER

In computer vision, maximally stable extremal regions (MSER) are used as a method of blob detection in images. This technique was proposed by Matas et al. to find correspondences between image elements from two images with different viewpoints. This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms.

3.5 Scale Invariant Feature Transform (SIFT)

Scale invariant feature transform (SIFT) [9] extracts and connects feature points in images which are invariant to image scale, rotation and changes in illumination [9]. Moreover, it provides distinctive descriptors that can find the correspondences between features in different images. Because of all these advantages, it is very suitable for estimating motion between images. Although SIFT has achieved remarkably success in video stabilization, it suffers costly computation, especially for low-end video cameras and cell phones. This motivates an intensive search for replacements with lower computational cost. Demonstrably, the best of these methods is speeded up robust features (SURF) [9].

3.6 Speed Up Robust Feature (SURF):

SURF is invariant to image rotation, scale, illumination and the overall robustness of the algorithm provides, in a substantial range, invariance to affine transformations. This algorithm is developed by Herbert Bay, TinneTuytelaars and Luc Van Gool. SURF algorithm can find correspondences between images based on keypoints within these images. When a large number of features are found, the algorithm is also invariant to occlusion. Occlusion happens when a part of an object is hidden behind another object. SURF outperforms SIFT in terms of result and computation time. Thus we choose SURF as our feature detector instead of other feature detector [8].

4. COMPARATIVE ANALYSIS

A comparative analysis of features extraction methods and The main characteristics of image features are shown in table 1 and table 2 respectively.

Fig 3(A): Corridor Image **Fig 3(B): Library Image**

(Image 1) (Image 2)

Table 1: Comparison Features Extraction Methods

Detector Type	Time(s) (Img-1)	Time(s) (Img-2)	Detected features Image1	Detected Features Image2
Harris Feature	0.351	0.292	314	1734
FAST Feature	0.07	0.22	329	3012
MSER	0.177	0.609	250	2112
SURF	**0.156**	**0.154**	**332**	**2546**

Table 2: The Main Charecteristic Of Image Features[9].

Features	Main Attributes	Advantages/ Limitations
Harris	–Strong Corner & Salient Points	**Advantage:** – Desirable detector in terms of detection & repeatabilty rate **Limitation:** – Cost is required more computation time & not variant to large scale change
ORB	–Binary valued and compact feature – Robust to lighting, blur, and perspective distortion	**Limitation:** Low degree of scale invariance
FAST	–Corner detector method – Faster than Other corner detector method	**Advantage:** – Very Fast –High quality feature detection **Limitation:** – Not robust to high levels noise
MSER	– Ridge based feature detection	**Advantage:** – Resulting regions
	method –"extremal" refers to the property that all pixels inside the MSER have either higher or lower intensity than all the pixels on its outer boundary. – Most efficient among the affine Invariant feature detector.	are accurately localized as compared to other blob detector.
SIFT	–Keypoint detector & descriptor –Keypoint detection is done by scale space representation of original image –Robust matching across a range of noise addition, affine distortion, and change in illumination/ viewpoints – Invariant to scale and rotation	**Advantage:** – Good recall rates(accuracy) – Relatively efficient as compared to other – Better error tolerance with fewer matches **Limitation**: – Still quite slow – Generally does not work well with lighting changes and blur
SURF	– Robust local keypoint detector & descriptors – Keypoint found by fast hessian detector that based on approximation of Hessian matrix for given image point – Relies on integral images to reduce the computation cost – RST invariant	**Advantage:** – Faster than SIFT – Good at handling blurring & image rotation **Limitation:** – Poor approximation of key point orientations – Poor at handling viewpoint change & illumination change

5. PROPOSED WORK

Here we surveyed that some work has been done thus far on video stabilization but majority algorithms fail in efficient reconstruction of boundary regions and thus results in poor reconstruction or stabilization of the video. Here we try to focus on the above mentioned problem through this work.
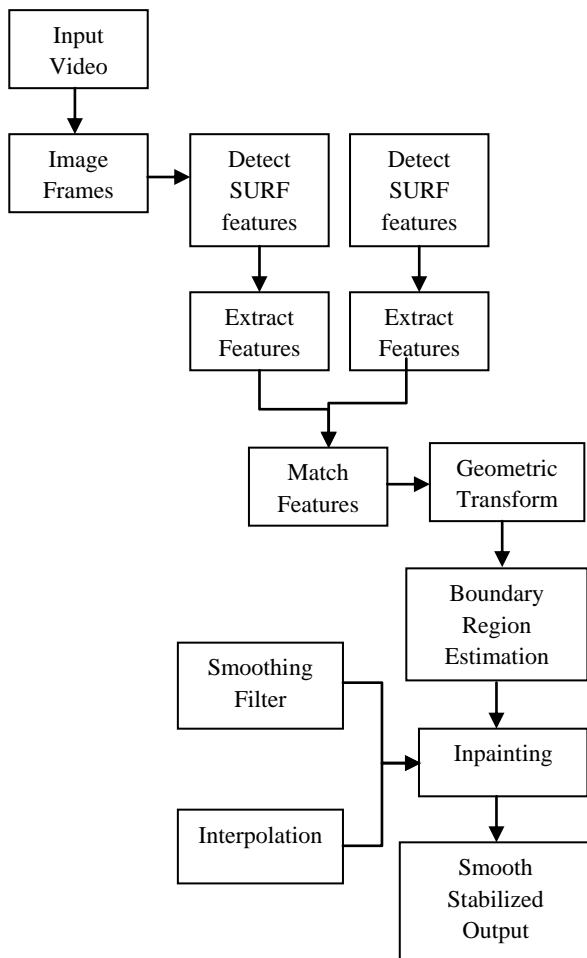
Fig 4: Flow Chart of Proposed Work

6. CONCLUSION

With having being done the literature review and a lot of work on the topic it was believed and now verified the same that the proposed method is better as compared to existing state of art methods for video stabilization as far as stability and boundary region reconstruction is concerned and also the comparison of the same are shown considering some standard parameters like PSNR, SSIM, MSE etc. of original and stabilized frames.

7. EXPERIMENTAL ANALYSIS

Here as compared to other state of art methods that uses other feature descriptors as compared to surf are slow and seems to be odd as when applied to videos having comparatively high resolutions. We propose an algorithm which uses SURF features and is very much supposed to give Stabilized output video with comparatively smooth boundaries. For boundaries reconstruction we use a novel inpainting algorithm.

8. ACKNOWLEDGEMENT

9. REFERENCES

[1] Yu-Hsi Chen and Hsueh-Yi Sean Lin "Full-frame Video Stabilization via SIFT Feature Matching" 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

[2] Yang Zhang, Yuquan Leng, Xu He Member, "A Fast Video Stabilization Algorithm with Unexpected Motion Prediction Strategy" 2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM) July 7-11, 2015. Busan, Korea.

[3] Sebastiano Battiato, Giovanni Gallo, Giovanni Puglisi, Salvatore Scellato" SIFT Features Tracking for Video Stabilization" 14th International Conference on Image Analysis and Processing (ICIAP 2007).

[4] 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.

[5] 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).

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

[7] Ankita Kansal1and Prof. Gulshan Goyal "Performance Analysis Of SIFT and SURF Approaches For Video Stabilization" International Journal of Scientific & Engineering Research, Volume 6, Issue 7, July-2015.

[8] Xie Zheng, Cui Shaohui, Wang Gang, Li Jinlun "Video Stabilization System Based on Speeded-up Robust Features" International Industrial Informatics and Computer Engineering Conference (IIICEC 2015).

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

[10] Trupti P. Patel,Sandip R. Panchal" Corner Detection Techniques: An Introductory Survey" International Journal of Engineering Development and Research 2014.

[11] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-Preserving Warps for 3D Video Stabilization," ACM Transactions on Graphics (Proceedings of SIGGRAPH 2009), 2009.

[12] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, Vol. 60, No. 2, Nov. 2004.

[13] 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.

[14] Michael L. Gleicher and Feng Liu, "Re-Cinematography: Improving the Camera Dynamics of Casual Video," Proc. ACM Multimedia,Sep. 2007.
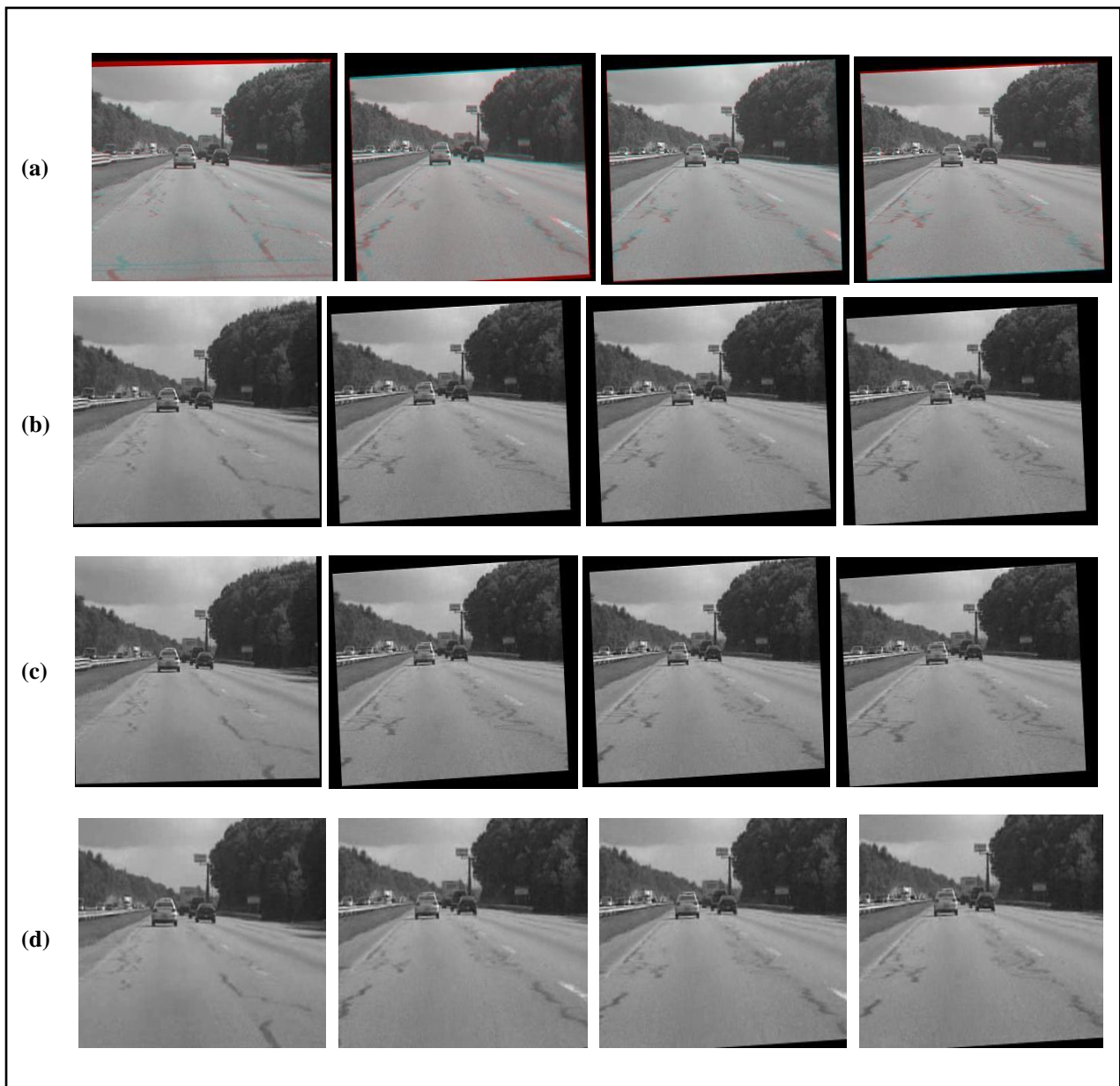
10. APPENDIX

Fig 5: Stabilize Outcomes using (a) Harris Feature (b) FAST Feature (c) MSER Feature (d) SURF