A Competitive Analysis of MP4 Video Stabilization

Rahul[1], Reema[2], Shabnam Kumari[3]

Department of CSE

[1]Mtech scholar, Sat Kabir Institute of Technology & Management, Bahadurgarh, Haryana, India

[2, 3]Assistant Professor,Sat Kabir Institute of Technology & Management, Bahadurgarh, Haryana, India

Abstract-*Stabilization is the process of estimating and compensating for background image motion occurring due to the ego-motion o camera. Video stabilization is a vital technique to remove visually distributing shaky or unstable motions from videos. The poor image quality of many video surveillance cameras effectively renders them useless for the purposes of identifying a person, a license plate, etc. Under certain conditions, however, it may be possible to combine multiple video frames for such identification tasks. To this end, we describe a simple and computationally efficient technique for video stabilization and enhancement. The efficacy of this technique is shown on synthetic and real video sequences.*

Keywords*-*Stabilization, Video Stabilization, Motion Estimation, CEVA, RANSAC.

I. INTRODUCTION

Many videos suffer from a shaking or moving camera and these camera movements tend to only distract the viewer from the subject matter. However, these movements can be fixed in software, in a process known as video stabilization. Many methods focus on point correspondences, where points from one frame of the video are matched with points in the next frame. If a transformation that aligns the points from one frame with their corresponding points in the other frame can be found, then the underlying video frames can be aligned using the same transformation. Points are matched using descriptors, which quantify characteristics of the points, so that points with the best matching descriptors are the most likely to be matching on the underlying frames. These descriptors are traditionally based on local features, which are the properties of the region of the frame surrounding a point.

II. RELATED WORK

Image stabilization is a key preprocessing step in video analysis and processing. In general stabilization is a kind of warping of video sequences, in which the image motion due to camera rotation or vibration is totally or partially removed. Most proposed algorithms compensate for all motions so the resultant background remains motionless. The motion models in [4] usually combine a pyramid structure to evaluate the motion vectors with an affine motion model to represent rotational and translational camera motions, and others [3] use a probabilistic model with a Kalman filter to reduce the motion noises and to obtain stabilized camera motions. Chang et al [1] use the optical flow between consecutive frames to estimate the camera motion by fitting a simplified affine motion model. Hansen et al [5] describe an image stabilization system, which uses a multi-resolution, iterative process to estimate the affine motion parameters between levels of Laplacian pyramid images. The parameters through a refinement process achieve the desired precision. This paper describes a simplified stabilization algorithm where an iterative process based on a coarse to fine fashion is used. The motion vectors are firstly estimated using the block matching technique, between two successive fields, and then the dense motion field is estimated using the motion vectors and the Horn-Schunck algorithm. By fitting an affine motion model, the motion parameters are computed and the currently stabilized ith video frame is based on the previously stabilized frame or the original ith frame. The ambiguity between image motion caused by 3D rotation and that caused by 3D translation (per frame) is solved by analyzing the direction of motion vectors [9, 10] and their standard deviation.

III. PROBLEM STATEMENT

Methods that rely on being able to distinguish local features will fail on pathological videos where many regions are near identical. These methods try to identify points in a frame by the image patch surrounding them, and thereby ignore significant useful information such as the overall structure of the frame. Our method replaces the local feature descriptors with the global contextual descriptor used by the Belonged. This descriptor, here after known as the BMP descriptor, looks at the point's context within the frame. It can therefore take into account broad structural aspects and work well even on data where local features fail to distinguish points. The properties of the BMP descriptor are desirable in aligning videos where other methods fail, such as low frame rate videos where the content changes significantly between frames. The BMP descriptor relies mainly on a global context rather than pixel intensities, so it avoids making assumptions about the data set maintaining similar intensities throughout. This assumption is inherent in matching descriptors of local features, and also pervades alternative methods such as Lucas Kanade feature tracking [6] and other optical flow methods[7,8] that try to identify regions moving between frames based on their appearance. The features from the Scale Invariant Feature Transform (SIFT) have also been used for video stabilization [11] due to being invariant under image scaling, rotation, and illumination; however, these features are complicated and expensive to compute, and the SIFT features still rely on the feature tracked remaining visually similar throughout in terms of intensity changes. The BMP descriptor's reliance on contextual information about the points allows it to overcome these other methods in challenging situations where the video frames change not just by changing illumination and movement, but also by changes in the addition and removal of textural information used by these other methods. To our knowledge, the application of the BMP descriptor to video stabilization is novel, although it has already been applied to matching points across images of a scene from multiple angles for the purpose of filling in regions of missing information in one angle using the information seen in the matching region from another angle [12]. We have examined the efficacy of the BMP descriptor for video stabilization and compared this method to a method using local feature descriptors and RAndom SAmple Consensus (RANSAC) to select the best matches [13].

3.1. RANSAC

A video frame is passed to a corner detector, which returns a collection of points at locations that appear to be the meeting point of two edges or transitions in the image. Each point is then given a local feature descriptor equal to a small image patch under the point. This process is done on each frame in the video. Then, starting at the first frame, the points in the frame are compared with the points in the next frame by comparing the descriptors of the points. Many potential matches of a point in the earlier frame to a point in the later frame are made. In all but the most trivial of examples, significant portions of these matches will be incorrect and severely hinder the accuracy of the method, so the matches are pared down by RANSAC, which takes random samples of point matches and removes the matches that seem to not agree with the majority of matches. This hopefully results in a set of matches that are consistent in the transformation between frames that they suggest. After this process is performed for each pair of consecutive frames in the video, the matches are used to compute transformations to align each video frame to the preceding one and thereby stabilize the entire video. This existing method works well with general videos [15], but we have identified two data sets that exhibit properties that lead to poor results and sometimes outright failure. However, the BMP descriptor is well suited to handle even these pathological cases. The particulars of these data sets will be expanded upon in the following section. The BMP descriptor is invariant under translation, and robust to small affine transformations [14]; which is applicable because we can reasonably assume that successive frames differ only by a small transformation. However, when the frame rate is low or when trying to align non-successive frames, the misalignment may be too large. We extend the BMP descriptor to be rotationally invariant such that it is still applicable in cases of frame rotation.

IV. PROPOSED ALGORITHM

The proposed algorithm computes a threshold of each block of an image adaptively based on the scatter of regions of change (ROC) using the local change adaptive thresholding and averages all thresholds for image blocks to obtain the global threshold based on entropy values. To enhance the segmentation results, the results thus obtained from ROC thresholding section are verified and compensated by considering the information of the pixels belonging to objects in the previous frame. After deciding the window size entropy of the window is computed from the gray level distribution of the window and background subtraction (non ROC region) is done after calculation of entropy.

We propose entropy based adaptive window selection scheme to determine the block/ window size in background segmentation process. Here the threshold value for a particular window is obtained by Otsu's thresholding scheme. To enhance the segmentation results, the results thus obtained from ROC segmentation are verified and compensated by considering the information of the pixels belonging to objects in the previous frame. This is represented asWhere, R is a matrix having the same size of the frame, s is the element number in the matrix and rs is the value of the VOP at location s. If a pixel is found to have rs = 1, it is a part of the moving object of the previous frame; otherwise it belongs to the background of the previous frame. Based on this information, Thresholding in Background (Non ROC) is modified as follows: if it belongs to a moving object part in the previous frame and its label obtained by temporal segmentation is the same as one of the corresponding pixels in the previous frame, the pixel is marked as the foreground area in the current frame else as a background.

4.1. Harris Corner Detection

The main goal of this step is to correct the distortion between the two frames by finding a transformation that will be done by applying an object system which returns affine transform [3]. The input for this stage should supply the object with a set of point correspondences between the two frames [4]. Firstly, the wanted points from the two chosen frames have to be identified followed by selecting the common correspondence between the frames. At this point, the candidate points for each frame are identified but to make sure that these points will have corresponding points in the second frame, it is necessary to find points around salient image features, like corners. Thus, Corner Detector System Object is used to find corner values using Harris Corner Detection which is one of the fastest algorithms to find corner values

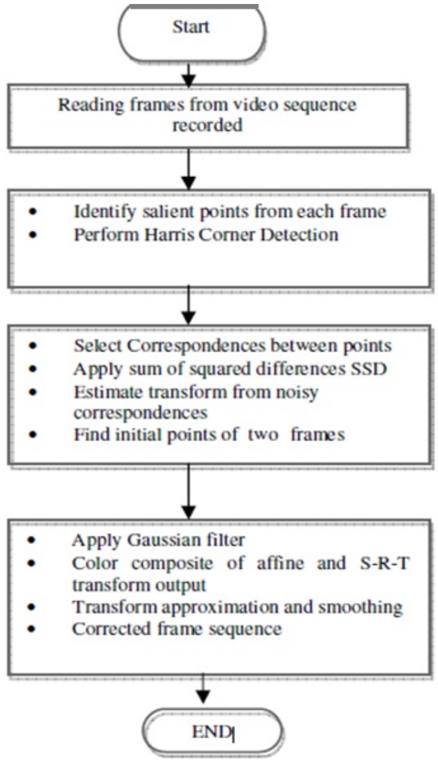

Figure 4.1: proposed algorithm

V. CONCLUSION

We have described a simple and computationally efficient technique for video stabilization and enhancement. The motion between video frames is modeled as a global affine transform whose parameters are estimated using standard differential motion techniques. A temporal mean or median filter is then applied to this stabilized video to yield a single high quality frame. We have shown the effectiveness of this technique on both synthetically generated and real video. This technique should prove useful in enhancing the quality of low-grade video surveillance cameras. Demand is on the rise for video cameras on moving platforms. Smartphones, wearable devices, cars, and drones are all increasingly employing video cameras with higher resolution and higher frames rates. In all of these cases, the captured video tends to suffer from shaky global motion and rolling shutter distortion, making stabilization a necessity. Integrating an embedded video stabilization solution into the imaging pipeline of a product adds significant value to the customer. It improves the overall video quality and, at the same time, enables better video compression and more robust object recognition for higher-level computer vision tasks.

REFERENCES

[1] M.M.Chane. kM.Tekalo. andM.I.Sezan. “Simultaneous motion estimation and segmentation,” IEEE Transaction on image Processing, VoL6, No. 9, pp. 1326-1333, Sept. 1997.

[2] J.Wangand E.Adelson, ‘‘ Representing moving images with laver.” IEEE Transaction on Image Processing. Vo1.3. NoS.DD. 6i5-638, Sept. 1994.

[3] P.H.S.Torr, RSzeliski, and P.Anandan,”An integrated Bayesian approach to layer extraction from image sequence,” IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 23, No.3, pp. 297-303, Mar. 2001.

[4] D. Wang, “Unsupervised video segmentation based on watersheds and temporal tracking,” IEEE Transaction on Circuits and Systems for Video Technology, Vol. 8, No.5, Sept. 1998

[5] C.Gu and M.C.Lee, ”Semiautomatic segmentation and tracking of semantic video objects,” IEEE Transaction on Circuits and Systems for Video Technology, Vo1.8, No.5, Sept 1998.

[6] Hiraiwa, A., Accurate estimation of optical flow for fully automated tracking of moving-objects within video streams, Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEEInternational Symposium

[7] Takaya, K., Tracking a video object with the active contour (snake) predicted by the optical flow, Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference

[8] Ping Gao, Moving object detection based on kirsch operator combined with Optical Flow, Image Analysis and Signal Processing (IASP), 2010 International Conference

[9] Khan, Z.H., Joint particle filters and multi-mode anisotropic mean shift for robust tracking of video objects with partitioned areas, Image Processing (ICIP), 2009 16th IEEE International Conference

[10] YuanyuanJia, Real-time integrated multi-object detection and tracking in video sequences using detection and mean shift based particle filters, Web Society (SWS), 2010 IEEE 2nd Symposium

[11] Sung-Mo Park, Object tracking in MPEG compressed video using mean-shift algorithm, Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia.Proceedings of the 2003 Joint Conference of the Fourth International Conference

[12] Yilmaz, A., Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale an Orientation Selection, Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference

[13] E. Y. Kim, S. H. Park and H. J. Kim, “A Genetic Algorithm-based segmentation of Random Field Modeled images,” IEEE Signal processing letters, vol. 7, No. 11, pp. 301-303, Nov. 2000.

[14] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 6, No. 6, pp. 721-741, Nov.1984.

[15] E. Y. Kim and S. H. Park, “Automatic video Segmentation using genetic algorithms,” Pattern Recognition Letters, vol. 27, No. 11, pp. 1252-1265, Aug. 2006.

[16] G. K. Wu and T. R. Reed, “Image sequence processing using spatiotemporal segmentation,” IEEE Trans. on circuits and systems for video Technology, vol. 9, No. 5, pp. 798-807, Aug. 1999.

[17] S. W. Hwang, E. Y. Kim, S. H. Park and H. J. Kim, “Object Extraction and Tracking using Genetic Algorithms,” Proc. of International Conference on Image Processing, Thessaloniki, Greece, vol.2, pp.383-386, Oct. 2001.

[18] Yilmaz, A., Javed, O., Shah, M., “Object tracking: A survey”, ACM Computing Surveys, Vol. 38,No. 4, Article No. 13 (Dec. 2006), 45 pages, doi: 10.1145/1177352.1177355, 2006.

[19] T. Bernier, J.-A. Landry, "A New Method for Representing and Matching Shapes of Natural Objects", Pattern Recognition 36, 2003, pp. 1711-1723.

[20] H. Blum, “A Transformation for extracting new descriptors of Shape”, in: W. Whaten-Dunn(Ed.), MIT Press, Cambridge, 1967. pp. 362-380.

[21] Rajaguru, D., Moving object detection using draw pad, Advances in Engineering, Science and Management (ICAESM), 2012 International Conference

[22] HaihuiGao, A New Detection Algorithm of Moving Objects Based on Human Morphology, Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference

[23] Chih-Hsien Hsia, Improved directional lifting-based discrete wavelet transform for low resolutionmoving object detection, Image Processing (ICIP), 2012 19th IEEE International Conference

[24] Xie, Songyun, Fast detecting moving objects in moving background using ORB feature matching, Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference