International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

Survey paper for Different Video Stabilization Techniques

Dipali Umrikar [1], Sunil Tade[2]

[1]*P.G. Student, Department of Electronics and Telecommunication Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*

[2] *Associate Professor, Department of Electronics and Telecommunication Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India*

Abstract - *The disturbances, that occurs on the video are produced by translational and rotational movements and by the zoom of the camera and the local motion of objects. Vibrations and shocks always occur on the camera while platform is moving. Other effects such as wind may also cause distortion. These causes degradations on the quality of the video. Video stabilization is the technique of generating a stabilized video sequence, where image motion by the camera's undesirable shake. It removes those undesired motions while preserving the desired motions. Different strategies and calculations have been produced during recent years. This paper summarizes the three robust feature detection methods: Scale Invariant. Feature Transform (SIFT), Speeded Up Robust Feature (SURF) and Block Based method to analyze the result in video stabilization application. SIFT presents its stability in most situation although it is slow. SURF is faster as compared to SIFT. Block based method has simple calculations, high anti-noise capacity, good stability for video stabilization.*

***Key Words*: SIFT, Block matching, SURF, video stabilization algorithms, Motion Estimation, Motion Smoothing.**

1.INTRODUCTION

Video stabilization technique uses either hardware or software inside the digital camera to minimize the effects of camera shake or vibration. Camera blur is more pronounced when shooting in low-light conditions, when using a long zoom lens where the camera's shutter speed slower to allow lighter to reach the camera's sensor. Due to the slower

shutter speed, any vibration occurring with the camera is magnified and sometimes causing blurry photos. Sometimes the slightest movement of your hand or arm could cause a blur. As most of the cameras are hand-held, mounted on moving platforms or subjected vibrations, this is an important technique in present day digital cameras. The proposed methods work at the frame level by classifying the inter-frame camera motion patterns. Regardless of the way that an extensive measure of progress has been made in the past 30 years, super settling certifiable video progressions still remains an open issue. By far most of the past work acknowledge that the concealed development has a fundamental parametric edge, and moreover that the dark piece and upheaval levels are known. Regardless, in fact, the development of things and cameras can be subjective, the video may be dirtied with upheaval of darken level, and development cloud and point spread limits can provoke to a dark part. Along these lines, a suitable super assurance system should at the same time assess optical stream, bustle level and cloud partition despite reproducing the high-res plots. As each of these issues has been particularly analyzed in PC vision, it is typical to merge each one of these parts in a lone structure without making distorted assumptions. So, for ongoing usage we concentrated couple of more video adjustment calculation. This paper describes detail steps of video stabilization and provides performance analysis of various techniques.

2. RELATED WORK

This paper [1] presents a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale and rotation. This paper has also presented methods for using the keypoint object recognition. The main approach described approximate nearest neighbor lookup, a Haugh Transform for identifying clusters, least square pose determination and final verification. In this paper [2] the SIFT feature is applied

to estimate camera motion. The unwanted camera vibrations are separated with combination of Gaussian Kernel Filtering and Parabolic Fitting. The paper [5] summarizes the three feature detection methods- Scale Invariant Feature Transform (SIFT), Principal Component Analysis (PCA), Speeded Up Robust Features (SURF). The performance of these methods is compared for scale and illumination changes, rotation, affine transformations.

In this paper [3] authors show a novel approach for video stabilization based on Speeded Up Robust Feature (SURF) method. Global motion is estimated by Heuristic Modified Trellis Search method. Unwanted shaky motion is filtered out by Kalman filter. A Comparison is given to time taken to find out feature point by SIFT and SURF. In this paper [4] the authors have proposed a real-time video stabilization based on block matching method Kalman filter is used to remove shaking of video and protects the panning motion properly.

3. ALGORITHMS FOR VIDEO STABILIZATION

As of late we have saw that the market of hand-held camcorders has development quickly in prevalence. In any case, the recordings recovered from such gadgets are influenced by undesirable camera shakes and nerves, bringing about video quality misfortune. Thus, video adjustment is a system that is utilized to enhance the video quality by expelling the undesirable camera developments because of hand shaking and inadvertent camera shake. The point of video adjustment is to smooth the obscured video that is brought on by undesired development in camera. In the previous decades, various looks into have been done in the video adjustment field. For the most part, the procedure of video adjustment comprises of three noteworthy strides: movement estimation, movement smoothing, and video finishing.

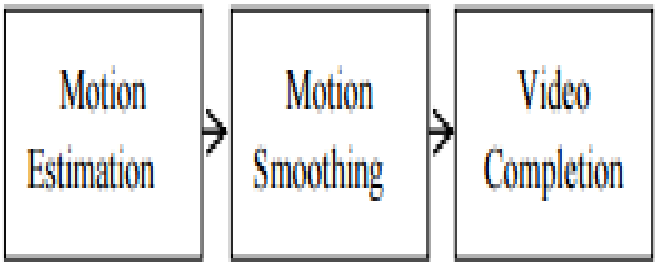

Fig.-1: General Video Stabilization Process

A. Scale Invariant Feature Transform (SIFT)

Scale invariant component change (SIFT) extricates and associates highlight focuses in pictures which are invariant to picture scale, revolution and changes in light. Also, it gives

unmistakable descriptors that can discover the correspondences between components in various pictures. In light of every one of these focal points, it is exceptionally appropriate for assessing movement between pictures. In spite of the fact that SIFT has made amazingly progress in video adjustment, it endures expensive calculation, particularly for low-end camcorders and phones. This rouses a concentrated look for replacing with lower computational cost. Evidently, the best of these strategies is accelerate hearty elements (SURF).

- **Scale Space Extrema Detection**:

A function, L(x, y, σ) is the scale space of an image which is generated by the convolution of Gaussian function, G(x, y, σ) and an input image, I(x, y):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

D which is computed by convolving the difference of two nearby scales separated by a constant scaling factor 'k' with an input image.

$$D(x, y, \sigma) = \big((G, x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) \quad (3)$$

- **Keypoint Localization**:

Keypoint selection from extrema can be done by rejecting the points along image edges and with low contrast and unstable over image variations. The feature points need to meet the below equation otherwise it is eliminated.

- **Orientation Assignment:**

To accomplish turn invariance, each keypoint is relegated an introduction

- **Keypoint Descriptor Generation:**

The estimations of introduction histogram, in both picture plane and scale space shape the descriptor. With exhibit of histograms and 8 introduction containers in every, results in component highlight vector.
SIFT features are used instead of extracting common corners or boundaries which always produce discreditable result. These features are affine invariant and non-sensitive to scale changes and illumination changes. Gaussian Filtering combined with Parabolic Fitting method is applied to estimating intentional motion. Finally, Dynamic programming (DP) method is proposed to fill up the missing area [2]. The primary contribution of this paper [2] are:
- tracking the SIFT features to estimate the global motion.
-use of DP to fill up missing areas.

Fig. -2: Result of video stabilization [2]. First row: original input sequence: frame numbers are 5, 10, 15, 20. Second row is stabilized sequence with missing areas. Third row stabilized and mosaicing sequence

B. Speeded Up Robust Feature (SURF)

SURF is a hearty picture intrigue point finder and descriptor plot, initially displayed by Herbert Bay et al. in 2006. SURF descriptor is like the angle data removed by SIFT [4] and its variations, while portraying the appropriation of the force content inside the intrigue point neighborhood. SURF is said to have comparative execution to SIFT, while in the meantime being speedier. The critical speed pick up is because of the utilization of vital pictures, which definitely decrease the quantity of operations for straightforward box convolutions, autonomous of the picked scale.

- **Interest Point Localization**

The SURF detector is the determinant of Hessian for scale selection and based on the Hessian matrix for its good performance in accuracy. Given a point x = (x, y) in an image I, the Hassian matrix is

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (4)$$

Where, Lxx(x,) -Convolution of Gaussian second order derivative of image I at point x, $Lxy(x,\sigma)$ and $Lyy(x,\sigma)$.

- **Interest Point Descriptor:**

With the end goal of getting invariance to picture revolution SURF first uses the Haar wavelet reactions in x and y bearing to register a reproducible introduction, then builds a square district adjusted to the chose introduction and concentrates the SURF descriptor from it. The Haar wavelet can be immediately ascertained by fundamental pictures. The windows can be part up in 4*4 sub-areas when the predominant introduction is assessed and incorporated into the intrigue focuses. The basic power example of every sub locale can be portrayed by a vector.

$$V = \left(\sum dx\ , \sum dy, \sum |dx|, \sum |dy|\right) \quad (5)$$

here dx, stand for the Haar wavelet response in horizontal direction and dy, is Haar wavelet response in vertical direction. And $|dx|$ and $|dy|$ are the absolute values of responses.

- **Matching by Nearest neighbor**

The separation proportion Matching of descriptors should be possible by closest neighbour remove proportion method (NNDR). In this strategy, the Euclidean separation between the descriptor of the component guide which is toward be coordinated, and its coordinating hopefuls are discovered. Authors have described the algorithm as follows [2].

Step1: Finding the SURF feature and descriptors. The SURF points and associated descriptor of each frame are found out.

Step2: Finding the correspondences between feature points are established by finding the set of matched pair or descriptors by NNDR method.
Step3: Track the feature points through the frames by Modified Trellis Search Method [3].
Step4: Rejection of feature points with local motion.

To determine whether the Kth feature point is of a moving object or not.

$$\psi(k) = \frac{1}{n}\sum_{i=1}^{n-1}\left(pos_k(i+1) - pos_k(i)\right) \quad (6)$$

Where $pos_k(i)$ is the position of k^{th} feature point in the i^{th} frame and n is number of frames in which that particular point exists. Motion filtering is done by Kalman filter. In this paper [3] authors found out SURF points and associated descriptors of each frame. They proved that SURF gives high repeatability than existing method. A comparison is given for the time taken to find out SURF and SIFT method [3]. Experimental result shows the unstable and stabilized video sequences [3].

Table -1

Frame No	SURF		SIFT	
	Feature points	Time	Feature points	Time
Frame 1	126	0.85 s	142	1.34 s
Frame 10	132	0.89 s	155	1.65 s
Frame 15	121	0.81 s	139	1.41 s
Frame 20	136	0.91 s	160	1.78 s

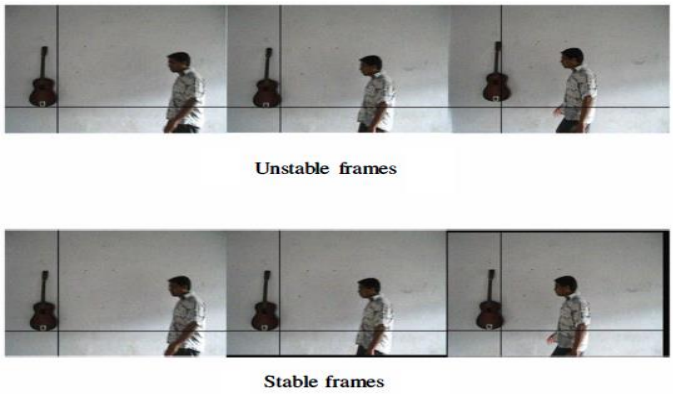

Fig.-3: Experimental result - Unstable and stable video sequence [3].

C. Block based Method

This technique is the productive strategy for adjustment. In this approach, every video outline gets partitioned into macroblock (estimate: 16*16). Macroblocks of current edge and past edges are coordinated on the premise of certain square coordinating calculation. Fig 4 demonstrates ventures of square based movement estimation, for example, pick movement vector (x, y), select macroblock of size M X N, pick look go p, seek best coordinating piece. There are some cost capacities which are considered as criteria for coordinating obstructs, these are given beneath:

a) The Mean Absolute Difference (MAD) is a model used to figure out which piece ought to be utilized. Normally, the lower the MAD the better the match and piece with the base MAD is picked. MSE which computes Mean Squared Error.

$$MSE = \frac{1}{N^2}\sum_{i=0}^{N-1}(C_{ij} - R_{ij})^2 \quad (7)$$

$$MAD = \frac{1}{N^2}\sum_{i=0}^{N-1}|C_{ij-}R_{ij}| \quad (8)$$

Where N is width or height of macroblock
Cij is Pixels compared in current macroblock
Ri is Pixels compared in previous macroblock

b) Sum of Absolute Difference (SAD) is given by:

$$SAD = \sum_{t=o}^{N-1}\sum_{j=0}^{M-1}|fcurrent(j,i) - fref(j + V_{x,i} + V_y| \quad (9)$$

Where, N= Height of block, M= Width of block, i=Index of horizontal direct, j= Index of vertical direction.
Vx, Vy = Motion vectors of reference block, fcurrent (x,y)= Pixel intensity at current block, f_{ref} (x,y) = Pixel intensity of reference block.

c) PSNR (Peak Signal to Noise Ratio) is calculated by

$$PSNR = 10\log_{10}[\frac{(I_{max})^2}{MSE} \quad (10)$$

Where $Imax$ = Max value of intensity of pixel.

$$E(u,v) = \sum_{x,y} w(x,y)\,[I(x+u, y+v) + I(x,y)]^2 \quad (11)$$

Window function is Gaussian window which gives weights to pixels. The function E(u, v) is to be maximized for corner detection.
The main idea is based on hierarchical block matching [4]. For an incoming video sequence, prominent key blocks are picked throughout the whole image. Global motion vector (GMV). Then local motion vector is calculated by fast hierarchical block matching to generate global motion vector through offline model. After word, the global motion vector is sent to Kalman filter. Finally, the stable output video is obtained. The whole image is divided into M*N big regions. The representative blocks of size m*m which contains most of the frame features are picked to decrease the calculations. Sum of absolute difference SAD is used to measure the abundance of each block because more difference inside the block generates better machine result.

$$S_{tx=} \sum_{h=1}^{m}\sum_{k=1}^{n}|I_t(x+h, y+k) - I_t(x+h, y+k-1)| \quad (12)$$

$$S_{ty} = \sum_{h=1}^{m}\sum_{k=1}^{n}|I_t(x+h, y+k) - I_t(x+h, y+k-1)| \quad (13)$$

$$S_t = S_{tx} + S_{ty} \quad (14)$$

Where I_t is the grey value of blocks in the region, S_{tx} and S_{ty} are the abundance of X axis and Y axis respectively, S_t represents abundance of block calculated.

The block with biggest abundance as a key block of the region chosen. The move vectors of each keyblocks in each layer are calculated using SAD and transported to the next layer after multiplied by two until the bottom layer.

$$f_t^{(l+1)}(i,j) = f_t^{l}(2i, 2j), \quad l = 0,1 \ldots \ldots N \quad (15)$$

Where $f_t^{l}(i,j)$ is the pixel at co-ordinate (i, j) in the l-layer of frame t.

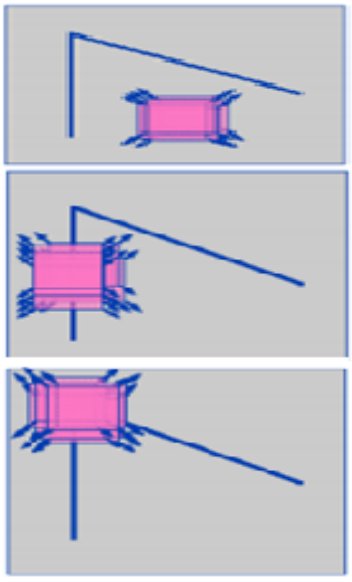

Fig -4: (a)flat region no change in all direction, (b) edge no change in edge direction, (c)corner region: significant region in all direction

The affine model is then applied with the generalized least square method to estimate global motion vector (GMV).

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = S \begin{bmatrix} \cos\theta & sin\theta \\ sin\theta & \cos\theta \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \quad (16)$$

Where S represents the zoom factor of camera and θ for rotation angle while dx and dy are the horizontal and vertical GMV respectively, (x, y) and (x', y') are matched key block center co-ordinates in current and reference frame.
Kalman filter is used to discard jitter motion. Authors have used median of last L frames as adaptive parameters.

$$M_{ed}(M_v) = \max\left(\sum_{i=k-l}^{k-1} M_{vi}\right) + min\left(\sum_{i=k-l}^{k-l} M_{vi}\right) \quad (17)$$

Where M_{vi} is the specific motion vector (i.e. SMV and GMV) in the i[th] reference frame, k is the current number.

$$R_k = \frac{k}{M_{ed}(SMV) + \alpha \cdot M_{ed}(GMV)} \quad (18)$$

Where k is the constant value, R_k represents the covariance of the observation noise, α factors the rate of response to the start of panning empirically set as 0.5.

Fig. -5: Experimental result [4]. First row shows unstabilized video sequence. Second row shows stabilized video sequence.

4. CONCLUSIONS

In this paper, a fast and robust video stabilization algorithm is derived. A method to perform the video stabilization has been the detection of SIFT feature and it has been good stability but high computational density. SURF is faster as compared to SIFT. Block based method has simple and fast calculations, high anti-noise capacity, good stability for video stabilization.

REFERENCES

[1] David G. Lowe, 'Distinctive Image Feature Scale-Invariant Keypoints', International Journal of Computer Vision, DOI 10.1023/B- VLSI. 0000029664.99615.94, 60(2), 91-110, 2004.

[2] Rong Hu, Rongjie, I-fan Shen and Wenbin Chen, 'Video Stabilization using Scale Invariant Features', 11th International Conference Information Visualization (IV'07), 2007 IEEE.

[3] Binoy Pinto and P.R. Anurenjan, 'Video Stabilization using Speeded Up Robust Feature', International Conference on communications and Signal Processing 2011. Pages 527-531 DOI: 10.1109/ICCSP 2011.5739378.

[4] Lengyi Li, Xiaohong Ma and Zheng Zhao, 'Real-time video stabilization based on fast block matching and improved Kalman filter', Fifth International Conference on Intelligent Control and Information processing, Pages: 394-397 DOI: 10.11.9/ ICICIP.2014.7010285, 2014.

[5] Luo Juan, Ounong Gwun, 'A comparison of SIFT, PCA-SIFT and SURF', International Journal of Image Processing (IJIP), Pages 143-152, Volume 3, Issue 4, 2009.