ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SURVEY ON VIDEO STABILIZATION TECHNIQUES

Patel Amisha [*], Ms.Hetal Vala
Master of computer engineering, Parul Institute of Engineering and Technology, India

ABSTRACT

In computer vision, video stabilization is an image processing method to remove visually distributing shaky or unstable motions from videos. This image disturbance may be due to the handshake of the camera holder or platform vibrations in case of surveillance cameras. In order to remove these unsteady motions, video stabilization contains three major step Motion estimation, Motion compensation or Motion smoothing, and Image warping. There are mainly four categories of video stabilization viz., optical stabilization, electronic stabilization, mechanical stabilization, digital stabilization. Digital video stabilization techniques make use only of information drawn from already captured footage and do not need any additional hardware or knowledge about camera physical motion, making it inexpensive and relatively easy to use. Video stabilization is used for astrophotography, tracking targets in military, earth motion etc. Many approaches of video stabilization have been proposed. In this paper provide an analysis of different techniques used for digital video stabilization.

KEYWORDS: Video Stabilization, Motion Estimation, Motion smoothing, Image warping

INTRODUCTION

Vision system plays important roles in many intelligent applications, such as transportation system, security systems, monitoring systems. Cameras may be installed on building or held by a people. The videos taken from hand held mobile cameras suffer from different undesired and slow motions like track, boom or pan, these affect the quality of output video significantly. Stabilization is achieved by synthesizing the new stabilized video sequence; by estimating and removing the undesired inter frame motion between the successive frames. Generally the inter frame motion in mobile videos are slow and smooth. [2]

RELATED WORK

Most previous video stabilization methods follows the same framework and on improving components.
Video stabilization techniques can be broadly classified as mechanical stabilization, optical stabilization and image post processing stabilization. Mechanical stabilization systems based on vibration feedback through sensors like gyros accelerometers etc. have been developed in the early stage of camcorders [1]. Optical image stabilization, which has been developed after mechanical image stabilization, employs a prism or moveable lens assembly that variably adjusts the path length of the light as it travels through the camera's lens system. It is not suited for small camera modules embedded in mobile phones due to lack of compactness and also due to the associated cost. The digital image stabilization tries to smooth and compensate the undesired motion by means of digital video processing. In the image post processing algorithm, there are typically three major stages constituting a video stabilization process viz. camera motion estimation, motion smoothing or motion compensation, and image warping.[2]

Motion Estimation: Video stabilization is achieved by first estimating the interframe motion of adjacent frames. The interframe motion describes the image motion which is also called global motion. By using different motion estimation techniques it is possible to estimate object motion or camera motion observed in video sequence. Object motion defined as local motion of the scene, and camera motion is defined as the global motion. The motion estimation technique can be classified as feature based approaches or direct pixel based approaches. Feature based approach is faster than direct pixel based approach.[3]

Motion Smoothing: The goal of motion compensation is to remove high-frequency jitters from the estimated camera motion. It is the

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

component that most video stabilization algorithm attempt to improve and many methods have been proposed, such as particle filter, kalman filter, gaussian filter.[3]

Image warping: Image warping wraps current frame according to the smoothed motion parameters and generates the stabilized sequence.[3]

Vedio stabilization

The digital image stabilization tries to smooth and compensate the undesired motion by means of digital video processing. In the image post processing algorithm, there are typically three major stages constituting a video stabilization process viz. camera motion estimation, motion smoothing or motion compensation, and image warping. Various techniques have been proposed to reduce the computational complexity and to improve the accuracy of the motion estimation. The global motion estimation can either be achieved by feature based approach or pixel based approach. Feature based methods are generally faster than pixel based methods but they are more prone to local effects and there efficiency depends upon the feature point's selection. Hence they have limited performance for the unintentional motion. Essentially features are relevant points in the image which may be easily tracked between different images. Feature tracking estimates the motion of the frame by selecting features from the previous frame and finding them in the current one, evaluating how these points moved between frames. It is clear that features should be accurately and efficiently tracked and then coupled among different frames without errors: techniques use corners, edges, regions, textures and intersections. [2]

METHODS OF VIDEO SYNOPSIS

1. **SIFT(scale invariant feature transform)**
2. **SURF(Speeded up robust feature)**
3. **FAST(Feature from accelerated segment test)**
4. **BRIEF(Binary robust independent elementary feature)**
5. **ORB(Oriented FAST and rotated BRIEF)**

1. Scale Invariant feature Transform(SIFT)[4]

SIFT keypoints of objects are first extracted from a set of reference imagesand stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidian distance of their feature vectors. From the full set of matches, subsets of keypoints that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches.

The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract and allow for correct object identification with low probability of mismatch. They are relatively easy to match against a (large) database of local features but however the high dimensionality can be an issue. SIFT algorithm consists of four major stages: scale-space extrema detection, keypoint localization, orientation assignment and keypoint descriptor.

Scale-space extrema detection: This is the stage where the interest points, which are called keypoints in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images is taken. Keypoints are then taken as maxima/minima of the Difference of gaussian (DOG) that occur at multiple scales. DOG image D(x, y, σ) is given by:

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

Where $L(x, y, \sigma)$ is the convolution of the original image I(x, y) with theguassian blur $G(x, y, k\sigma)$ at scale kσ.

***Keypoint Localization*:** Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

The initial implementation of this approach (Lowe, 1999) simply located keypoints at the location and scale of the central sample point. However, recently Brown has developed a method (Brown and Lowe, 2002) for fitting a 3D quadratic function to the local sample points to determine the interpolated location of the maximum, and his experiments showed that this provides a substantial improvement to matching and stability. His approach uses the Taylor expansion

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

(up to the quadratic terms) of the scale-space function, D(x, y,σ), shifted so that the origin is at the sample point:

Eliminating Edge response:
For stability, it is not sufficient to reject keypoints with low contrast. The difference-of-Gaussian function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise.
A poorly defined peak in the difference-of-Gaussian function will have a large principalcurvature across the edge but a small one in the perpendicular direction. The principal curvaturescan be computed from a 2x2 Hessian matrix, H, computed at the location and scale of the keypoint:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2)$$

The derivatives are estimated by taking differences of neighboring sample points.

***Orientation Histogram*:** By assigning a consistent orientation to each keypoint based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation. This approach contrasts with the orientation invariant descriptors of Schmid and Mohr (1997), in which each image property is based on a rotationally invariant measure. The disadvantage of that approach is that it limits the descriptors that can be used and discards image information by not requiring all measures to be based on a consistent rotation.
The scale of the keypoint is used to select the Gaussian smoothed image, L, with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample, L(x, y), at this scale, the gradient magnitude, and m(x, y), and orientation, θ(x, y), are precomputed using pixel differences:

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2} \quad (3)$$

$$\theta(x,y) = \tan^{-1}\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x-1,y))} \quad (4)$$

SURF[6]

SURF[6] is a robust image interest point detector and descriptor scheme, first presented by Herbert Bay et al. in 2006. SURF descriptor is similar to the gradient information extracted by SIFT [4] and its variants, when describing the distribution of the intensity content within the interest point neighborhood. SURF is said to have similar performance to SIFT, while at the same time being faster. The important speed gain is due to the use of integral images, which drastically reduce the number of operations for simple box convolutions, independent of the chosen scale.

A. Interest Point Localization

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

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (5)$$

Where, $L_{xx}(x,\sigma)$ is the convolution of the gaussian second order derivative with the image I at point x, and similarly for $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$.[5]

B. Interest Point Descriptor

For the purpose of obtaining invariance to image rotation SURF first uses the Haar wavelet responses in x and y direction to compute a reproducible orientation, then constructs a square region aligned to the selected orientation and extracts the SURF descriptor from it. The Haarwavelate can be quickly calculated by integral images. The windows can be split up in 4*4 sub-regions when the dominant orientation is estimated and included in the interest points. The underlying intensity pattern of each sub region can be described by a vector

$$V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \quad (6)$$

where d_x, stand for the Haar wavelet response in horizontal direction and d_y, is Har wavelate response in vertical direction. And $|d_x|$ and $|d_y|$ are the absolute values of responses.[5]

C. Maching by Nearest neighbor distance ratio

Matching of descriptors can be done by nearest neighbor distance ratio method(NNDR)[7]. In this method,the Euclidean distance between the descriptor of the feature point which is to be matched, and its matching candidates are found out. If the ratio of first two minimum distances is less

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

than a threshold T, the descriptor that corresponds to the numerator is taken as a match.[5]

2. Feature Accerelated segment Test(FAST)[8]

FAST is an algorithm proposed originally by Rosten and Drummond [8] for identifying
interest points in an image. An interest point in an image is a pixel which has a well-defined position and can be robustly detected. Interest points have high local information content and they should be ideally repeatable between different images. Interest point detection has applications in image matching, object recognition, tracking etc. The reason behind the work of the FAST algorithm was to develop an interest point detector for use in real time frame rate applications.

The algorithm is explained below:

1. Select a pixel „p" in the image. Assume the intensity of this pixel to be I_p. This is the pixel which is to be identified as an interest point or not.
2. Set a threshold intensity value T, (say 20% of the pixel under test).
3. Consider a circle of 16 pixels surrounding the pixel p. (This is a Bresenham circle [4] of radius 3.)
4. "N" contiguous pixels out of the 16 need to be either above or below I_p by the valueT, if the pixel needs to be detected as an interest point. (The authors have used N =12 in the first version of the algorithm)
5. To make the algorithm fast, first compare the intensity of pixels 1, 5, 9 and 13 of the circle with IP. As evident from the figure above, at least three of these four pixels should satisfy the threshold criterion so that the interest point will exist.
6. If at least three of the four pixel values - I1, I5, I9 I13 are notabove or below I_p + T, then P is not an interest point (corner). In this case reject the pixel p as a possible interest point. Else if at least three of the pixels are above or below I_p+ T, then check for all 16 pixels and check if 12 contiguous pixels fall in the criterion.
7. Repeat the procedure for all the pixels in the image.[9]

3. Binary robust independent elementary feature(BRIEF)[10]

The BRIEF descriptor [10] is a recent feature descriptor that uses simple binary tests between pixels in a smoothed image patch. Its performance is similar to SIFT in many respects, including robustness to lighting, blur, and perspective distortion. However, it is very sensitive to in-plane rotation. It is a bit string description of an image patch constructed from a set of binary intensity tests. It provides a shortcut to find the binary strings directly without finding descriptors.
Consider a smoothed image patch, p.
The binary test τ is defined by:

$$\tau(P;x,y) = \begin{cases} 1 : p(x) < p(y) \\ 0 : p(x) \geq p(y) \end{cases}, \quad (7)$$

Where p(x) is the intensity of p at a point x.
The feature is defined as a vector of n binary tests:

$$f_n(p) = \sum_{1 \leq = i \leq = n} 2^{i-1}\ \tau(P;x_i,y_i) \quad (8)$$

One of the best performance a gaussian distribution around the center of the patch is used in BRIEF method. Vector length n is chosen as 256.

4. Oriented FAST and Rotated BRIEF (ORB) [11]

ORB modifies the FAST [8] detector to detect key points by adding a fast and accurate orientation component, and uses the rotated BRIEF [10] descriptor. Corner detection using FAST is carried out and that result in N points that are stored based on the Harris measure. A pyramid of the image is constructed, and key points are detected on every level of the pyramid. Detected corner intensity is assumed to have an offset from its center. This offset representation, as a vector, is used to compute orientation. Images are smoothened with the 31×31 pixel patch. Orientation of each pixel patch is then used to steer the BRIEF [10] descriptor to obtain rotational invariance.
The Table 1 shows the comparison between different methods for video synopsis.

Table 1. Comparison table of methods

Method	Advantage	Limitation
SIFT	• It extract distinctive feature from images that can be invariant to image scale and rotation.	• slow and not good at illuminatio n changes
SURF	• SURF is fast and has good performance as the same as SIFT.	• not stable to rotation and illuminatio n changes
FAST	• FAST is a corner detection method, which could be	• It is not robust to

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

	used to extract feature point and develop an interest point detector for use in real time frame rate applications.	high levels noise.
BRIEF	• BRIEF is a recent feature descriptor that uses simple binary tests between pixels in a smoothed image patch. • construction and matching much faster than other methods • higher recognition rates, as long as invariance to large in-plane rotations • real-time matching performance achieved with limited computational time	• It is not designed to be rotationally invariant.
ORB	• computationally efficient respect to SIFT, less affected by image noise, almost two orders of magnitude faster than SIFT and SURF.	• ORB is not designed to be scale invariant.

CONCLUSION

In this paper a variety of feature extraction techniques for motion estimation in video stabilization such as scale invariant feature transform, speeded up robust feature, Feature acceleration segment test, binary robust independent elementary feature and oriented FAST and robust BRIEF are studied. For eachtechnique a detailed explanation of the techniques can be given which are used for global motion estimation step of video stabilization. From this survey, a number of shortcomings and limitations were highlighted in each and everytechnique.

ACKNOWLEDGEMENTS

I am very grateful and would like to thank my guide and teacher Asst.Prof. Hetal Vala for her advice and continued support. Without her it would not have been possible for me to complete this paper. I would like to thank all my friends, colleague and classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious.

REFERENCES

[1] PareshRawat, 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.

[2] Multimedia, "Use Image Stabiliza. For Gyroscopic Stabilizer", [online], URLhttp://www.websiteoptimization.com/speed/tweak/stabilizer. Access 13-January-2009.

[3] AndriusAucinasHomerton College "Report on Digital video stabilization" Computer science Tripos, July 2011.

[4] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 2004.

[5] Binoy Pinto Dept. of Electronics and Communication College of Engineering Trivandrum "Video Stabilization using Speeded Up Robust Features" 978-1-4244-9799-7/11/2011 IEEE 527.

[6] Herbert Bay , Andreas Ess , TinneTuytelaars , Luc Van Gool "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding 110 (2008) 346–359.

[7] Vilar F. da CamaraNeto and Mario Fernando M. Campos, "An Improved Methodology for Image Feature Matching", XXII Brazilian Symposium on Computer Graphics and Image Processing, 2009.

[8] E. Rosten and T. Drummond, "Machine learning for high speed corner detection," in 9th Euproean Conference on Computer Vision, vol. 1, 2006, pp. 430–443.

[9] Deepak GeethaViswanathan "Features from Accelerated Segment Test (FAST)" 2011.

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. "Brief: Binary robust independent elementary features." In European Conference on Computer Vision, 2010.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF", 2011 IEEE International Conference on Computer Vision, pp. 2564-2571, Nov. 2011