International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Real Time Video Stabilization using PLK Tracking Algorithm

Sunil Sankol S M[1], Praveenkumar B S[2], Vanishree K Rao[3]

[1]Assistant Professor, Electrical Engineering Dept, MVJCE, Bengaluru-67, sankolsunil[at]gmail.com
[2]Assistant Professor, Mechanical Engineering Dept, DBIT, Bengaluru-74, praveenbsha[at]gmail.com
[3]Assistant Professor, Electrical Engineering Dept, MVJCE, Bengaluru-67, vanishreekrao6[at]gmail.com

Abstract: ***Video Stabilization is a technique to improve video quality. Hence, the unwanted motion fidget removal and enrichment technology which aims in removing troublesome trembling motion from videos. Its impact slopes upward rapidly with increasing popularity of handheld cameras and the cameras mounted on moving platforms comprehensive video stabilization becomes essential. The proposed approach allows for Video Stabilization beyond the conventional filtering of the camera paths and some methods are based on mechanical devices such as gyro sensor to detect the camera motion and then shift the image sensor to compensate vibration. To overcome an imperfection, the algorithm deals with estimation of camera motion path by optical flow method using Pyramidal Lucas Kanade (PLK) Feature Tracking algorithm. Motion estimation is calculated by applying optical flow of common coverage areas. In a practical the proposed techniques has been applied in a various real time objects tracking as a applied in various real time objects tracking as a pre-processing stage. Experimental observations show that the method can perform real time and provide good performance.***

Keywords: Stabilization, Pyramidal Lucas Kanade (PLK), Descritization

1. Introduction

Video stabilization is a video processing technique to enhance the quality of input video by removing the undesired camera motions. While digital still photography has advanced to the stage where most amateurs can easily get high-quality pictures, the quality gap between professional and amateur-level video remains remarkably wide. One of the biggest components of this gap is camera motion. Most camera motions in casual videos shot are hand-held, yielding videos that are difficult to watch. There are various approaches used for stabilizing the captured videos. Most of the existing methods are either very complex or does not perform well for slow and smooth motion of hand held mobile videos. Hence it is desired to synthesize a new stabilized video sequence, by removing the undesired motion between the successive frames of the hand held video devices. Various 2D and 3D motion models are used for the motion estimation and stabilization. Here various motion models, motion estimation methods and the smoothening techniques are explained. It also describes the direct pixel based and feature based methods of estimating the inter frame motion. Some of the results of the differential motion estimation are also presented. Finally it closes with an open discussion of research problems in the area of motion estimation and stabilization."A stabilized video is defined as a motionless video where the camera motion is completely removed"[1]. With the fast development of camera phones, there has been a dramatic increase in the amateur videos shot over the past decades.

However, people find such videos tough to observe, mainly due to the excessive amount of shake and undirected camera motions in the footage. Rattled camera motion and platform vibrations can be difficult to avoid when using handheld cameras, which will generate unstable video images. It is necessary to preserve the intentional camera motion while removing the undesired motion due to unsteady platform. However to ensure the visual quality of the whole video, video stabilization has a particular emphasis on the accuracy and robustness over long image sequences.

2. Literature Survey

Current Stabilization approaches employ key-point feature tracking and linear motion estimation in the form of 2D transformations, or use Structure from Motion (Sfm) to estimate the original camera path. From this original shaky camera path a new smooth camera path is estimated by either smoothening the linear motion models to suppress high frequency jitter, or fitting linear camera path augmented with smooth changes in velocity to avoid sudden jerks. The final step is synthesizing the stabilized video using the transformations obtained in smooth camera path estimation many methods just keep central parts of the original frames to achieve better visual quality.

Yasuyuki Matsushita et al. [2], defined video stabilization as "Motionless video where the camera motion is completely removed". They proposed a practical and robust approach of video stabilization comprising the methods of estimating camera Motion, motion smoothing and image warping which produces full-frame stabilized videos with good visual quality.

Zihan Zhou et al. [3], reviewed with a new image deformation technique called Content-Preserving Warping (CPW) which has been successfully employed to produce the state-of-the-art video stabilization resulting in many challenging cases. The key insight of CPW is that the true image deformation due to viewpoint change can be well approximated by a carefully constructed warp using a set of sparsely constructed 3D points only. However, since CPW solely relies on the tracked feature points to guide the warping it works poorly in large texture-less regions such as ground and building interiors.

International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Yu-Shuen Wang et al. [4], focused on handling parallax for video Stabilization and presents a robust and efficient technique that works on general videos. It achieves high-quality camera motion on videos where 3D reconstruction is difficult or long feature trajectories are not available. They represented each trajectory as a Bezier curve and maintain the spatial relations between trajectories by preserving the original offsets of neighboring curves. The technique formulates stabilization as a spatial-temporal optimization problem that finds smooth feature trajectories and avoids visual distortion.

3. Estimating Frame pair Transform

For motion estimation features are tracked using Pyramidal Lucas-Kanade. However, robustness demands good outlier rejection. For dynamic video analysis global outlier rejection is insufficient whereas the short baseline between adjacent video frames makes fundamental matrix based outlier rejection unstable. So to overcome this local outlier rejection is employed by discretizing features into a grid of 50x50 pixels, applying RANSAC[5] within each grid cell to estimate a translation mode, and only retaining those matches that agree with the estimated model up to a threshold distance. Grid based approach is used as it is faster. Subsequently, several 2D linear motion models (translation, similarity and affine) are fitted to the tracked features. Once the camera path is computed as a set of linear motion models, the optimal camera path is fit according to the framework.

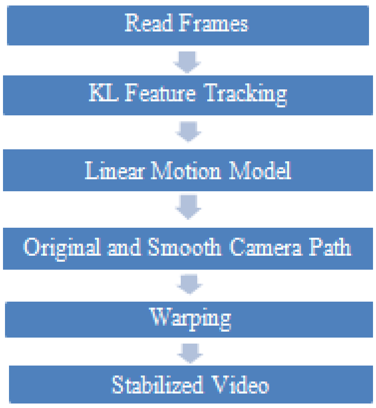

Figure 1: Methodology of Video Stabilization

Figure 1 shows the methodology of video stabilization where in first and foremost step is to calculate the width and the height of the video and number of frames present in the whole video following which the frames are extracted from the video. The sample video taken is of 360x640 which is of 30 frames per second (fps).

Figure 2: Feature point detection

Figure 2 shows the feature point detection indicated by green dots which is nothing but the edge in an image. The number of counts of feature can be changed in the program so hence to get the correspondence between the consecutive images and can be carried out either by computing the motion or by corresponding points.

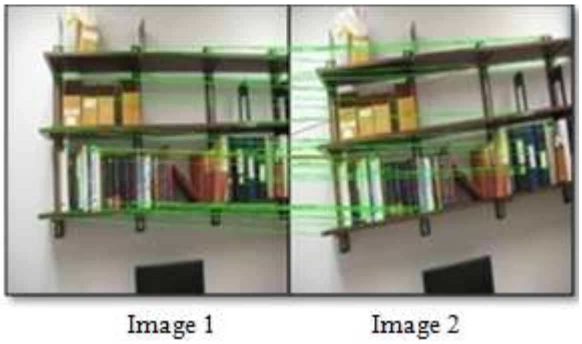

Figure 3: Feature point tracking between the Images.

The LK Tracker takes the input as pair of images and returns the list of points in Image1 and Image2 and there correspondences as shown in Figure3. The green line indicates the good feature points which are tracked from the Image1 to Image2 and the red lines are the bad feature points which are to be removed by applying the RANSAC algorithm. The bad feature points are the outliers and are removed by setting up the threshold distance to < 2 Pixels and also the number of iterations. Hence only the inliers are retained. For Motion Estimation, the features points in a frame (i.e. pair of images) are obtained and tracked by optical flow method using the pyramidal LK algorithm [6]. However, robustness demands good outlier rejection.

Computing the 2D linear motion models (translation, Similarity and affine) to the tracked feature is obtained by taking the affine transformation for every image i.e. F_t. Specifically if the video be a sequence of images

$$I_1, I_2 \ldots In$$

where each frame pair (I_{t-1}, I_t) is associated with a linear motion model $F_t(x)$ modeling the motion of feature points x from I_{t-1}, I_t.For example if there occurs no translation and no rotation then the matrix obtained is shown as below and vice-versa. Hence for series of images in a video of transform are obtained. The Discretized camera path C_t defined at each frame I_t, C_t is iteratively computed by the matrix multiplication, which is expressed as

$$C_t = F_1 \text{ x } F_2 \text{ x } F_3 \ldots\ldots\ldots F_t$$

Apply the LP [7] solver and find parameters for B_t such that the total objective function F(x) is minimized which is expressed as follows

$$F(x) = w_1 f_1(x) + w_2 f_2(x) + w_3 f_3(x)$$

where x is the number of parameters in case of affine transform $f_1(x)$, $f_2(x)$ and $f_3(x)$ are the three objective functions to be minimized and w_1, w_2 and w_3 are the respective weights (constants) for each of the objective functions, whose values indicates the relative importance of one objective function relative to the other.

The block of LP solver takes input as the coefficient of matrices, the RHS coefficients and lower bounds and upper bounds if any and returns the values of unknown parameters and the minimum value of the f(x) for which the objective function is said to be minimized i.e. fmin(x).Warp refers to the warping of the frames according to the update transform that is Bt, which transforms a crop window originally centered in the frame rectangle (Figure 4). In general we wish to limit how much B_t can deviate from the original path to preserve the intent of the original video.

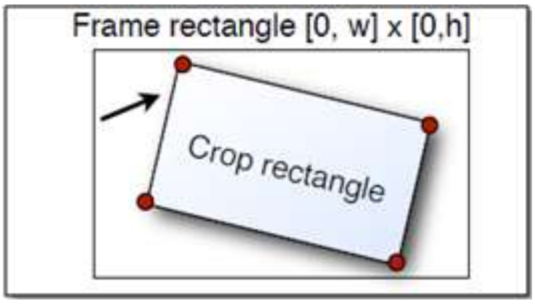

Figure 4: Cropped Rectangle applied for the video.

The crop rectangle is the window of the frame in the video which is in our case 75% to 80%.It is considered as an iterative process where in increase in size of cropped window results in the out of bound regions in the output video. Hence it is set with in this limit.

4. Results and Discussions

The videos are experimented with different feature detecting technique where the results are tabulated below. Corner detection is an approach used within computer vision systems to extract certain kinds of features and infer the contents of an image.

To detect corners:

1) Define and set up your corner detector using the constructor.
2) Call the step method with the input image, I, the corner detector object, cornerDetector, and any optional properties.

The corner Detector function is an object which finds corners in a gray scale image. It returns corner locations as a matrix of [x y] coordinates. The object finds corners in an image using the Harris method. The maximum corners have to been mentioned as a variable in the program i.e., 100.

Harris corner detection (Harris & Stephens)

Figure 5: Screenshot of Harris Corner Detector in the Video Frame.

Figure 5 shows the corners using the Harris corner detector where the points with high variance are chosen, as the computation is carried out for every pixel it consumes high computational time.

FAST Corner Detection Method

Figure 6: Screenshot of FAST Feature Detection in the Video Frame

Figure 6 shows the SURF [8] feature detector for a gray scaled image which is nothing but a 2D image. The features are indicated by green positive points encircled with green circle.

SURF Feature Detector: Detect SURF features and return SURF Points object.

Figure 7: Screenshot of SURF Feature Detectionin the Video Frame

Figure 7 shows the SURF feature detector for a gray scaled image which is nothing but a 2D image. The features are indicated by green positive points encircled with green circle.
Time taken for each of the feature detection instruction for 80% crop window for resolution 360x640 is given in the below table.

Table: Time consumption.

Feature Detector	Time (sec)
Harris corner detection (Harris & Stephens)	1.51
Local intensity comparison (Rosten & Drummond)	1.37
Minimum eigenvalue (Shi &Tomasi)	1.39

Above table 1 shows the comparison between all the three types of feature detection techniques. As Local intensity takes block of pixels and compares with the consecutive frame it is faster and hence takes less time than the other two Techniques.

Figure 8: Screenshot of Input video and Stabilized video

The above frame of Figure 8 shows the screenshot of the input jittery video and frame below is the output stabilized video. The crop window of size 80% of the full screen is shown with a black lined box which follows the smooth trajectory giving a stable output.

5. Conclusion

The proposed algorithm achieves good results with the trimming technique making it computationally feasible and applicable to variety of videos. Gridding of the video into blocks for the removal of the outliers helps to keep the good outliers rejection than to compute with each and every pixel. As the input video provides large set of data inherently corrupted by high noise levels, feature-based method, with Tomasi-Shi and Lucas-Kanade algorithms helps to quickly select input data to estimate inter-frame motion. Video Stabilization can be used in the field of real time object tracking or targeting as a pre-processing stage or module on the machines which travels faster at uneven grounds where there is no compromise in the visual information received. As the crop window size is fixed for a video to be stabilized, the future scope lies in overcoming this limitation.

References

[1] Grundmann M., Kwatra V, and Essa I, "Auto-directed Video Stabilization with Robust l1 Optimal Camera Paths, " in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 225–232.
[2] Yasuyuki Matsushita, EyalOfek, WeinaGe, Xiaoou Tang and Heung-YeungShum, "Full-Frame Video Stabilization with Motion Inpainting ", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, NO. 7, July 2006, 1150-1161.
[3] Zihan Zhou, University of Illinois at Urbana-Champaign, Hailin Jin, Adobe Systems Inc. and Yi Ma Yi Ma, Microsoft Research Asia, "Plane-Based Content-Preserving Warps for Video Stabilization", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
[4] Yu-Shuen Wang, Feng Liu, Member, IEEE, Pu-Sheng Hsu, and Tong-Yee Lee, Senior Member, IEEE, " Spatially and Temporally Optimized Video Stabilization", IEEE Transactions on Visualization And Computer Graphics, August 2013, Vol. 19, No.8.
[5] Konstantinos G. Derpanis,"Overview of the RANSAC Algorithm", kosta[at]cs.yorku.ca, May 13, 2010, Version 1.2.
[6] JeanYvesBouguet"Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm", Intel Corporation Microprocessor Research Labs.
[7] Singiresu S. Rao,"Engineering Optimization Theory and Practice", John Wiley & Sons, Inc Fourth Edition, 2009.
[8] Herbert Bay, TinneTuytelaars, and Luc Van Gool, "SURF: Speededup Robust Features", http://www.robots.ox.ac.uk/~vgg/research/affine.