VIDEO STABILIZATION

Sebastiano Battiato, University of Catania, Catania, Italy
Rastislav Lukac, Epson Canada Ltd., Toronto, Ontario, Canada

Synonyms: *Camera shake compensation.*

Definition: *Algorithms used to improve video quality by removing unwanted camera shakes and jitters due to hand jiggling and unintentional camera panning.*

Introduction

Video stabilization technology is used to avoid visual quality loss by reducing unwanted shakes and jitters of an image/video capturing device without influencing moving objects or intentional camera panning [1]. This is particularly essential in handheld imaging devices which are more affected by shakes due to their smaller size. Unstable images are typically caused by undesired hand jiggling and intentional camera panning whereas unwanted position fluctuations of camera result in unstable image sequences. Using video stabilization techniques ensures high visual quality and stable video footages even in non-optimal conditions. Ideally, imaging devices are equipped with mechanical means which physically avoid camera shakes, or they employ sophisticated optical or electronic means to first sense camera movement and then compensate for it by acting on the lens or on the image sensor. To reduce the cost of imaging devices, desired stabilization effects can be obtained through digital video stabilization which uses information drawn from footage images, with no need for additional knowledge about camera physical motion [2]. Digital video stabilization can be implemented in both real-time and offline systems. Applying stabilization in video recording helps reduce the bit rate of data compression because the performance of motion estimation usually significantly improves after stabilizing the successive frames.

Hardware Stabilization

As already noted, there exist three hardware approaches, that is, optical, electronic, and mechanical, to achieving stabilization effects.

In optical stabilization [3], vibrations are compensated by varying the optical path to the sensor using a floating lens element moved orthogonally to the optical axis of the lens. Vibrations are revealed by sensors and then mechanically controlled lenses instantly compensate the jitter with a correction movement before visual data is recorded. Thus, the system response is synchronized with the vibration. Since no manipulation is done on visual data, optical stabilization preserves image quality. Unfortunately, high cost of optical stabilization devices prevents from including them in low-end digital cameras.

Electronic stabilization uses an electronic system to control the stabilization process [5]. If the system detects through its sensors a camera shake, it responds by slightly moving the image so that it virtually remains in the same position on the image sensor. This movement is obtained by readdressing the area of the image sensor which is read by the capturing chip. Since the used area is small, image motion induces blur and graininess with consequent image degradation. This issue can be solved using oversized sensors or by digitally zooming the image; however, both these approaches produce some resolution loss.

In mechanical stabilization, camera motion is detected by gyroscopes. The gyroscopic wheels, occupying opposed axes to each other, spin with high speed and physically resist camera vibrations, acting like an invisible tripod. Once the camera motion is detected, the sensor is counter-moved to avoid vibrations and to obtain clear, steady images and jitterless panning effects.

Digital Stabilization

Unlike hardware stabilization solutions, digital stabilization systems operate on the captured image data [2],[4],[6]. The generic architecture of such systems is composed by three functional units, that is, motion estimation, motion filtering, and image composition. The motion estimation step compares the current frame with the previous one (in postcapture processing also with the next frame) to achieve the best approximation of the interframe motion vector. Then, this global motion vector is used in the motion filtering step to update the absolute motion vector which tracks camera movements frame by frame. Finally, the image composition step corrects the frame in order to obtain the stabilized sequence.

Digital stabilization often leads to visual quality degradations because image sensors are not fully used due to employed motion compensation. Basically, resulting video comprising of steady frames is captured using a smaller sensor area; thus reducing its resolution. Image borders are usually not shown in the final footage since they are used in digital stabilization to achieve the desired correction.

Motion Estimation

One of the main issues in digital stabilization is to estimate the global camera motion based only on the information contained in the video frames. The performance of a stabilization algorithm highly depends on the accuracy of estimated global motion vectors. Motion estimation in natural image sequences may be quite challenging; for example, due to the presence of large moving objects, low contrast areas and repeated patterns. Unfortunately, even if information gathered from frame content is essentially modeled in a two-dimensional space, camera motion occurs in a three-dimensional space. In spite of this, video stabilization can be accomplished by using simplified, two-dimensional motion models. This allows for relatively easy and fast processing of each individual frame and better performances even under real-time constraints.

Global motion estimation can be implemented using the methods and algorithms operating on some local assumptions. Optical flow based techniques [6] are computationally efficient; however, without multi-scale approaches they are not able to deal with large displacements. Feature-based algorithms extract features from video images and estimate interframe motion using locations of these features. There also exist techniques which combine concepts of feature extraction and robust filtering ([7], [8]). These methods have gained larger consensus for their good performances; unfortunately, calculating or extracting features is usually time consuming. Global intensity alignment approaches [9], which operate directly on image intensities to compute interframe transformation, are generally less sensitive to local image variations, but the problem can be their high computational cost. Finally, block matching-based techniques use different adaptive filters to refine initial motion estimation based on block local vectors [10]. These techniques generally provide good results and their computation complexity depends on complexity of the employed block matching step.

The whole set of local motion vectors (LMV) probably includes both incorrect and correct matches belonging to self-moving objects in the captured scene. Obviously, there exist some correct matches that represent real camera shakes. LMVs may not represent the exact motion vector of an image region if this region contains large motion objects, suffers from random noise, lacks high contrast patterns, or it is not exposed normally. These irregular LMVs should be bypassed and not used for final global motion estimation (Figure 1).

(a) (b)

Figure 1. Achieved motion vectors: (a) motion vectors obtained by a block matching algorithm operating in 16×16 blocks, and (b) residual vectors after the filtering process.

Alternatively only some specific regions of each frames can be used for LMVs computation, trying to identify specific (but fixed) areas where background regions are typically located (Figure 2).
Global motion between adjacent frames can be estimated with a two-dimensional linear model which usually provides a trade-off between effectiveness and complexity. This model describes interframe motion using four different parameters; namely, two shifts, one rotation angle and a zoom factor. The model associates a point (x_i, y_i) in frame I_n with a point (x_f , y_f) in frame I_{n+1} with the following transformation:

$$\begin{cases} x_f = x_i\lambda\cos\vartheta - y_i\lambda\sin\vartheta + T_x \\ y_f = x_i\lambda\sin\vartheta + y_i\lambda\cos\vartheta + T_y \end{cases}$$

where λ is the zoom parameter, θ the rotation angle, T_x and T_y respectively X-axis and Y-axis shifts. In order to estimate four transformation parameters, four different linear equations are required. Thus, only two couples of features allow for the system to find a solution. Since features can be affected by noise, it is useful to apply a linear least squares method on a set of redundant equations. The least squares method does not perform well when the set of features consists of a large portion of outliers. However, outliers can be identified, removed from the estimation process by filtering, thus resulting in higher accuracy. Many robust estimation techniques have been developed for computer vision applications [11], [12]; however, in order to obtain real-time performance a simplified version of iterative least squares methods [8], [13] can be used.

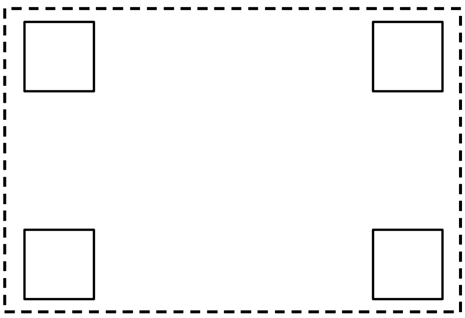

Figure 2. An example of regions used for LMV calculations.

Motion Filtering

It is quite common that a feature can move from a frame to the next one due to both camera shakes and intentional panning movements. In this situation, the corresponding local motion vectors may result in inaccurate motion estimation, since video stabilization should occur only on the camera jitters and should not compensate desired movements. To avoid this problem, motion filtering should be employed and estimated motion should be evaluated in order to recognize intentional movements. Several techniques, such as Kalman filtering [7], [14] and motion vector integration [15] have been proposed to correct translational and rotational jitters [10], [16] according to real systems constraints.

Image Composition

The final stage of an image stabilization system is the image composition process. With the estimated motion compensation vectors, the process stabilizes the image frame by adjusting the location offset of the sensor active area. In some cases, further processing (e.g., interpolation, inpainting, etc.) may be needed to fill up the missing image area and to deblur the images.
The stabilization process changes the location of the cropped window such that the relative positions of the objects belonging to the image background keep the same locations in the final frames. This compensation method is the most popular one in consumer digital cameras due to its relatively low computational complexity. The major drawback of this method is the reduced resolution of resulting frames due to trimming image areas. In addition to limited adjustable margins of sensor area, simple approaches usually cannot handle severe camera shakes as well as rotational types of camera vibrations. To avoid this problem, sophisticated image stabilization systems consider both translational and rotational motion models (Figure 3). Namely, unstable images are first translated and rotated to let the locations of the objects in the scene keep unchanged after stabilization.
Stabilized images often suffer from the trouble of some areas being trimmed. Filling up those missing image areas is called image or video inpainting [17]. Another issue in camera motion compensation is reducing image blurriness, which is also called motion deblur [1], [21]. Motion blur is caused by a moving scene point that spreads out several pixel locations during the exposure period of the sensor.

Evaluation

Automatic assessment of the video stabilization performance is important in order to tune and compare various techniques. In some cases the full stabilization cannot be achieved precisely. Furthermore, the shake compensation process itself often introduces some additional distortion (such as blurriness) in images instead of removing it.
Typically automatic criteria for image quality are based on a simple pixel-by-pixel comparison between the ground truth image and the reference image. Such pixel-wise approaches are often extended to video quality assessment, simply by comparing still images on a frame-by-frame basis [1]. A simple rough evaluation of a stabilization algorithm can be achieved by using the interframe transformation fidelity (ITF) measure [4] defined as follows:

$$ITF = \frac{1}{N_{frame} - 1} \sum_{k=1}^{N_{frame}-1} PSNR(k)$$

where $PSNR(k)$ is defined as

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

In the above expression, $MSE(k)$ is the mean square error between two consecutive frames, I_{max} is the maximum intensity value of a pixel, and N_{frame} is the video frames number. Typically, stabilized sequences have higher ITF values than their unstable versions.

Figure 3. Example of image composition: (top) four destabilized consecutive frames of a real sequence, and (bottom) stabilized sequence.

A more detailed evaluation can be obtained by measuring separately performances of the motion estimation and motion filtering modules in order to evaluate the amount of intentional camera movement (e.g., panning) erroneously removed [18], [19].

Finally, stabilized videos can exhibit some additional blurring to each compensated frame; especially when rotation and translational models are employed. A human observer often finds this blurring irritating, even if images were perfectly aligned. The amount of additional blurring introduced by the stabilization depends on the employed interpolation technique and other implementation details [1], [4], [17].

Still Photography

In digital photography, stabilization techniques are employed to enhance overall quality of images acquired in low-light conditions [20]. The exposure time needed to acquire the pixel values is long enough to generate some unpredictable motion blurring even in a single shot. In general, blurred (or degraded) images can be restored by considering the distortion operator, also called the point spread function (PSF). If the PSF is known, the original image may be restored by applying an image deconvolution algorithm [21]. However, in real cases, the PSF is usually unknown, thus requiring blind deconvolution approaches [22], [23]. When considering simple and easy to implement motion models, these approaches usually produce poor results in practice. Fortunately, there exist techniques which synthesize blur-free images by considering a short image sequence acquired with different exposure time. One possible way of stabilizing images leads to the fusion of a series of low-exposed images [24]. These images usually do not suffer from blur (because of low-exposure time) but have high noise level. In order to reduce computational and memory requirements, some approaches consider only a pair of images (instead of a number of images from the image sequence), typically consisting of a low-exposed image (no blur but high noise level) together with a normal exposed image (high blur degradation but low noise level) [20], [25], [26].

See: Motion estimation, motion blurring, video codec, exposure settings.

References

1. R.C. Gonzalez, R.E. Woods, "Digital Image Processing", Third Edition, Prenctice Hall, 2007.
2. F. Vella, A. Castorina, M. Mancuso, G. Messina, "Digital Image Stabilization by Adaptive Block Motion Vectors Filtering", IEEE Transactions on Consumer Electronics, vol. 48, issue 3, pp. 796–801, August 2002.
3. Canon Inc., Canon FAQ: What is Vari-Angle Prism?, http://www.canon.com/bctv/faq/vari.html , 2007.
4. L. Mercenaro G., Vernazza, C. Regazzoni, "Image Stabilization Algorithms for Video-Surveillance Application", in Proceedings of the IEEE International Conference of Image Processing, vol. 1, pp. 349-352, October 2001.
5. C. Morimoto, R. Chellappa, "Fast Electronic Digital Image Stabilization", in Proceedings of 13th International Conference on Pattern Recognition, vol. 3, pp. 284-288, August 1996.
6. J. Chang, W. Hu, M. Cheng, B. Chang, "Digital Image Translational and Rotational Motion Stabilization Using Optical Flow Technique", IEEE Transactions on Consumer Electronics, vol. 48, no. 1, pp.108-115, February 2002.
7. J. Yang, D. Schonfeld, C. Chen, M. Mohamed, "Online Video Stabilization Based on Particle Filters", in Proceeding of the IEEE International Conference on Image Processing, pp. 1545-1548, October 2006.
8. S. Battiato, G. Gallo, G. Puglisi, S. Scellato, "SIFT Features Tracking for Video Stabilization," in Proceeding of the IEEE International Conference on Image Analysis and Application, pp. 825–830, September 2007.
9. M. Tico, S. Alenius, M. Vehvilainen, "Method of Motion Estimation for Image Stabilization", in Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 277-280, May 2006.
10. S. Auberger, C. Miro, "Digital Video Stabilization Architecture for Low Cost Devices", in Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, pp. 266-271, September 2005.
11. M.A. Fischler, R.C. Bolles, "Random Sample Consensus: A Paradigm Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, vol. 24, issue 6, pp. 381–395, June 1981.
12. P. J. Huber, "Robust Statistical Procedures", SIAM, 1996.
13. A. Bjorck, "Numerical methods for least squares problems", SIAM, 1996.
14. S. Erturk, "Image Sequence Stabilisation Based on Kalman Filtering of Frame Positions", IEE Electronics Letters, vol. 37, no. 20, pp. 1217-1219, September 2001.
15. J. K. Paik, Y.C. Park, D.W. Kim, "An Adaptive Motion Decision System for Digital Image Stabilizer Based on Edge Pattern Matching", IEEE Transactions on Consumer Electronics, vol. 38, no. 3, pp. 607-616, August 1992.
16. M. Tico, M. Vehvilainen, "Constraint Translational and Rotational Motion Filtering for Video Stabilization", in Proceedings of the 13th European Signal Processing Conference, September 2005.
17. Y. Matsushita, E. Ofek, G. Weina, T. Xiaoou, S. Heung-Yeung, "Full-frame Video Stabilization with Motion Inpainting", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1150-1163, July 2006.
18. M. Niskanen, O. Silven, M. Tico, "Video Stabilization Performance Assessment", in Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 405-408, July 2006.
19. A. Engelsberg, G. Schmidt, "A Comparative Review of Digital Image Stabilising Algorithms for Mobile Video Communications", IEEE Transactions on Consumer Electronics, vol. 45, no. 3, pp. 591–597, August 1999.

20. Q.R. Razligh, N. Kehtarnavaz, “Image Blur Reduction For Cell-Phone Cameras Via Adaptive Tonal Correction”, in Proceedings of the IEEE International Conference on Image Processing, pp.113-116, October 2007.
21. P.A. Jansson, “Deconvolution of Image and Spectra”, Academic Press, 1997.
22. T.F. Chan, C.K. Wong, “Total Variation Blind Deconvolution”, IEEE Transactions on Image Processing, vol. 7, no. 3, pp. 370-375, March 1998.
23. Y.L. You, M. Kaveh, “A Regularization Approach to Joint Blur Identification and Image Restoration”, IEEE Transactions on Image Processing, vol. 5, no. 3, pp. 416-428, March 1996.
24. M. Tico, M. Vehvilainen, “Robust Image Fusion for Image Stabilization”, in Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 565-568, April 2007.
25. M. Tico, M. Vehvilainen, “Image Stabilization Based on Fusing the Visual Information in Differently Exposed Images”, in Proceedings of the IEEE International Conference on Image Processing, pp. 117-120, October 2007.
26. J. Jia, J. Sun, C. Tang, H. Shum, “Bayesian Correction of Image Intensity with Spatial Consideration”, in Proceedings of the European Conference on Computer Vision, pp. 342-354, May 2004.