Quantitative Analysis of Digital Image Stabilization

K. Dastagiri

M.Tech (DSCE), Student, Department of Electronics & Communication Engineering, RGMCET
Nandyal-518 501, Kurnool (dist), Andhra Pradesh, India

Abstract: ***Image stabilization has become a subject of significant interest and an active research field over the past years due to the wide use of digital imaging devices. The image stabilization process aims at removing irregular motion phenomena from image sequences in order to accomplish a compensated sequence that displays smooth camera movements. A variety of image processing applications requires motion-compensated image sequences as inputs. The unwanted positional vacillations of the video sequence will affect the visual quality and impede the subsequent processes for several applications. An innovative technique for digital image stabilization (DIS) based on the Hilbert–Huang transform (HHT) is studied. It exploits the basic features of the HHT in order to separate the local motion signal obtained from an image sequence into two different motion vectors. A variety of embedded systems equipped with a digital image sensor, such as handheld cameras, mobile phones, and robots, can produce image sequences with an observed motion caused by two different types of movements: the smooth camera motion (intentional) and the unwanted shaking motion (jitter). To verify the effectiveness of the proposed DIS method, several simulations were performed, and the results were compared with existing stabilization methods. An attempt has been made for estimate Mean square error (MSE), peak signal to noise ratio (PSNR) and structural similarity index (SSIM) to study the performance of HHT-DIS and compared with other existing techniques and showed that HHT-DIS outperforms the existing methods.***

Keywords: Digital image stabilization (DIS), Hilbert–Huang transform (HHT), image sequence processing, jitter motion designation.

1. Introduction

Image stabilization systems could be classified into four primary categories: optical, orthogonal transfer charge-coupled device (CCD), electronic and digital stabilizers. Optical image stabilizers are mechanisms used in still or video cameras that stabilize the recorded image by varying the optical path to the sensor [1]. The key element of all optical stabilization systems is that stabilization process is applied before the projected image is converted into digital information by the sensor. Moreover, often used in astronomy, an orthogonal transfer CCD shifts the image within the CCD itself while the image is being captured, based on analysis of the apparent motion of bright stars [2]. Electronic image stabilizers use motion gyroscope sensors attached to the camera in order to detect the camera movement. The image sequence is compensated using the opposite direction of the sensor's readings [3]. The digital image stabilization (DIS) is the process of removing the unwanted motion effects of a moving camera to produce a compensated image sequence by using image processing techniques. DIS methods outperform other image stabilization methods by means of hardware requirements and flexibility, since they are hardware independent and could be used in on-offline applications without any restrictions. DIS process applied after the image acquisition, consists of three main stages [4], [5], [7]; global motion estimation, jitter motion determination, and image warping. The first processing stage is dedicated to the determination of parameters that designates the entire camera motion, meaning that displacements between either blocks or features in subsequent frames are determined. On the contrary, local motion vectors (LMVs) are calculated within smaller frame regions during the process of motion estimation. Essentially, LMVs represent the offset of specific image regions between two consecutive frames. Thus, LMVs include both the intentional and the unwanted motion of the camera. This process dominates over the other two stages in terms of time complexity. In this paper an attempt has been made to estimate parameters like MSE, PSNR and SSIM for HHT-DIS method and comparison of these metrics have been carried out with existing techniques.

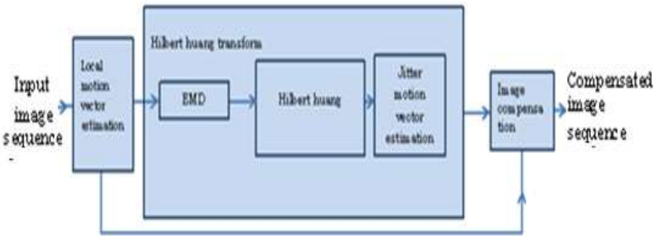


Figure 1: DIS method by HHT

2. Hilbert Huang Transform

Hilbert–Huang transforms (HHT) [6], [8] as a signal-processing tool that adaptively decomposes non stationary signals through the process of empirical mode decomposition (EMD) in to basic functions called intrinsic mode functions (IMFs). The HHT combines the EMD and the Hilbert spectral analysis (HSA); the HSA includes the Hilbert transform of each IMF generated by the EMD process. Fig.1. Shows block diagram representation of the DIS by HHT method [8].

First, a motion estimation method is applied in order to define the LMV of an image sequence. Subsequently, the estimated LMV is decomposed into a finite number of IMFs by applying the EMD process. Each IMF is transformed using Hilbert transformation in order to define the energy content of every decomposed signal. Depending on the estimated energies, the last IMF to be considered as jitter is designated. Thus, the summation of all IMFs with the lower indices up to the specified IMF from the Hilbert transform approximates the unwanted jitter motion. Finally, the image sequence is compensated according to the calculated sum in order to produce a stabilized image flow.

In this paper first we take a one video having length of small duration. Next a small amount of motion blur is added to the video. Motion blur is introduced to the video in terms of L & T. Here L is length of the blur in pixels; T is angle of the blur in degrees. After adding the motion blur to the video, remove the jitter motions from the video by using HHT - DIS method.

3. Simulation Results

To verify the effectiveness of the HHT-DIS method, several simulations were performed, and the results were compared with existing stabilization methods. In order to evaluate the performance of the method, four different image sequences were processed. The most widely used metric for such applications is the root MSE (RMSE) which was applied so that a quantitative comparison with the other related methods could be achieved and is calculated by

$$e_{rms} = \frac{1}{N}\sqrt{\sum_{n=1}^{N}(\bar{x}_n - x_n)^2 + (\bar{y}_n - y_n)^2} \quad (1)$$

Where N is the number of frames and (x_n, y_n) and $(\bar{x}_n, \bar{y}_n)$ are the optimal and the resulting camera motion, respectively. The under evaluation image sequences are of 200 or 250 frame length. For evaluation purposes, the intentional motion must be known in order to evaluate the difference between the optimal and the retrieved motion. The optimal motion of every image sequence was created by using a moving camera setup.

Peak signal-to-noise ratio, often abbreviated PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. PSNR is most commonly used to measure the quality of reconstruction of loss compression codec's (e.g., for image compression). Although a higher PSNR generally indicates that the reconstruction is of higher quality, in some cases it may not. One has to be extremely careful with the range of validity of this metric; it is only conclusively valid when it is used to compare results from the same codec (or codec type) and same content. PSNR is most easily defined via the mean squared error (*MSE*).

The PSNR is defined as:

$$\text{PSNR} = 10 \log_{10} (R^2/\text{MSE}) \quad (2)$$

Here R is the maximum fluctuation in the input image data. For example if the input image has a double precision floating point data type then R is 1. If it has an 8 bit unsigned integer data type R is 255, etc. When you try to compute the MSE between two identical images, the value will be zero and hence the PSNR will be undefined (division by zero). The main limitation of this metric is that it relies strictly on numeric comparison and does not actually take into account any level of biological factors of the human vision system such as the Structural Similarity Index (SSIM). For color images, the MSE is taken over all pixels values of each individual channel and is averaged with the number of color channels. The structural similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index is a full reference metric; in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference.

The difference with respect to other techniques mentioned previously such as MSE or PSNR is that these approaches estimate perceived errors; on the other hand, SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. The SSIM metric is calculated on various windows of an image. The measure between two windows x and y of common size N×N is:

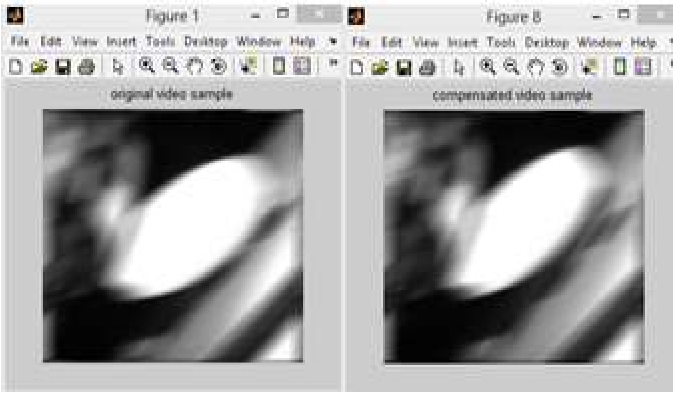
$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

With

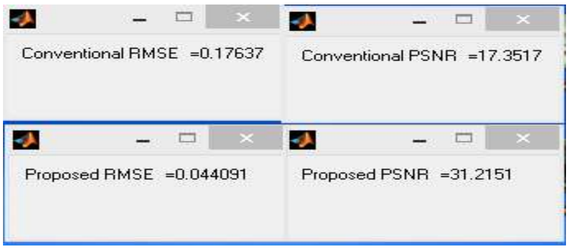
- μ_x the average of x;
- μ_y the average of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$);
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

In order to evaluate the image quality this formula is applied only on luma. The resultant SSIM index is a decimal value between -1 and 1, and value 1 is only reachable in the case of two identical sets of data. Typically it is calculated on window sizes of 8×8. The window can be displaced pixel-by-pixel on the image but the authors propose to use only a subgroup of the possible windows to reduce the complexity of the calculation.

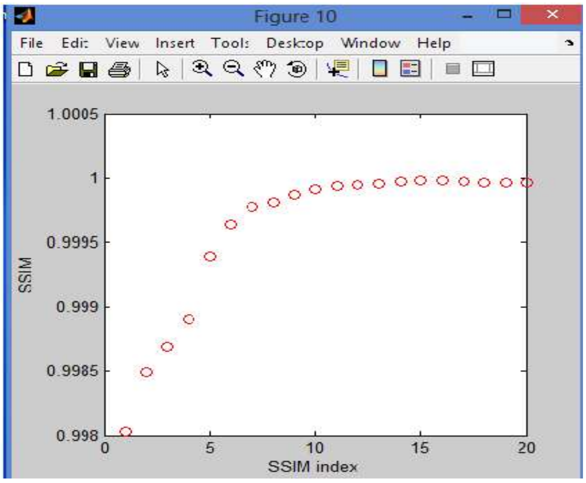
Take a video having length of small duration. Next a small amount of motion blur is added to the video. Motion blur is introduced to the video in terms of L & T. Here L is length of the blur in pixels; T is angle of the blur in degrees. After adding motion blur to the video remove the jitter motions from the video by using HHT method, and evaluate the parameters like MSE, PSNR and SSIM values to compare the performance of HHT-DIS method with existing techniques.



(a) (b)

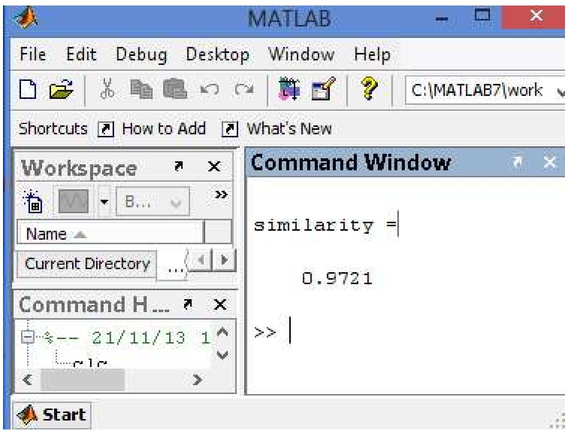


(c)



(d)

Figure 2: (a).original video sample having jitter motions (b) compensated sample (c) Computed RMSE & PSNR (d) SSIM values for different pixels.



(e)

Figure 2: (e) Over all SSIM value

Table 1: Following table shows the RMSE, PSNR & SSIM values obtained when the video have fixed angle of the blur in degrees (T) and variable length of the blur in pixels (L).

S. No	*Amount of Blur*	*Approach*	*Image Quality parameters*		
			RMSE	*PSNR*	*SSIM*
1	T = 15	Conventional	0.17637	17.3517	0.9721
	L = 30	HHT	0.04409	31.2151	
2	T = 15	Conventional	0.1762	17.3612	0.9754
	L = 32	HHT	0.4405	31.2244	
3	T = 15	Conventional	0.17595	17.3756	0.9616
	L = 34	HHT	0.04398	31.2393	
4	T = 15	Conventional	0.17548	17.4021	0.9721
	L = 36	HHT	0.04386	31.266	
5	T = 15	Conventional	0.17502	17.4284	0.9740
	L = 38	HHT	0.04375	31.2926	

Table 2: Following table shows the RMSE, PSNR and SSIM values obtained when the motion blur having fixed length of the blur in pixels (L) and variable angle of the blur in degrees (T).

S. No	Amount of Blur	Approach	Image Quality parameters		
			RMSE	PSNR	SSIM
1	L = 15	Conventional	0.18034	17.1293	0.9640
	T = 10	HHT	0.04507	30.9936	
2	L = 15	Conventional	0.18033	17.1296	0.9643
	T = 11	HHT	0.04507	30.9934	
3	L = 15	Conventional	0.18029	17.1316	0.9694
	T = 12	HHT	0.04507	30.9944	

4. Conclusion

In this paper, a novel DIS method based on the HHT has been studied. The method acquires the required data from an image sequence in order to achieve the definition of the two fundamental motions, the intentional and the trembled camera motion. These two motions are inserted in the image sequence due to both the camera movement and a no steady motion of the camera. Thus, image sequences captured with handheld cameras, mobile phones, or even with a camera of a mobile robot can be processed to produce a stabilized sequence with smoother frame transitions. A small amount of motion blur is added to the video in terms of L & T. For different values of L & T RMSE, PSNR and SSIM values are estimated for a given video and compared with existing methods. It has been observed HHT – DIS method is better compared to existing methods.

Reference

[1] B. Cardani, “Optical image stabilization for digital cameras,” *IEEE Control Syst. Mag.*, vol. 26, no. 2, pp. 21–22, Apr. 2006.

[2] B. E. Burke, R. K. Reich, E. D. Savoye, and J. L. Tonry, “An orthogonaltransfer CCD imager,” *IEEE Trans. Electron Devices*, vol. 41, no. 12,pp. 2482–2484, Dec. 1994.

[3] C. Morimoto and R. Chellappa, “Fast electronic digital image stabilization,”in *Proc. IEEE Int. Conf. Pattern Recogn.*, 1996, vol. 3, pp. 284–288.

[4] C.Wang, J.-H. Kim, K.-Y. Byun, J. Ni, and S.-J. Ko, “Robust digital image stabilization using the Kalman filter,” *IEEE Trans. Consum. Electron.*vol. 55, no. 1, pp. 6–14, Feb. 2009.

ISSN (Online): 2319-7064

[5] S.-C. Hsu, S.-F. Liang, and C.-T. Lin, "A robust digital image stabilization technique based on inverse triangle method and background detection," *IEEE Trans. Consum. Electron.*, vol. 51, no. 2, pp. 335–345, May 2005.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[7] Amanatiadis and I. Andreadis, "Digital image stabilization by independent component analysis," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 7,pp. 1755–1763, Jul. 2010.

[8] A Digital Image Stabilization Method Based on the Hilbert-Huang Transform,"IEEETrans.on Instrum. Meas., vol.61,no.9 September 2012

Author Profile

K. Dastagiri received his B-tech degree from R.G.M College of Engineering and Technology in Electronics and Instrumentation Engineering, and pursuing M-tech Degree in Rajeev Gandhi Memorial College of Engineering and Technology in the specialization of Digital Systems and Computer Electronics. His research interest is Quantitative Analysis of Digital Image Stabilization.