IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 5, Ver. III (Sep.- Oct. 2017), PP 23-30
www.iosrjournals.org

An Enhanced Video Stabilization Based On Emd Filtering And Spectral Analysis

*B. Sowbhagya
Assistant Professor, Ballari Institute of Technology and Management, Ballari, Karnataka, India.
*Corresponding Author: *B. Sowbhagya*

Abstract: *In this paper a new video stabilization technique is proposed based on the Hilbert Huang Transform (HHT). HHT is a Decomposes the video into Intrinsic Mode Functions (IMFs). The proposed video stabilizer deals with vertical or horizontal displacements meaning that translational motions are corrected. First, LMVs from a certain image region of the sequence are defined by applying a motion estimation algorithm. Second, the resulting signal is separated into a finite number of IMFs through the process of EMD. The proposed HHT-based DIS method can be beneficial in a variety of applications including handheld cameras and mobile phones as a post-processing method, as well as for stabilizing image sequences acquired from cameras mounted on vehicles.*

Keywords: *Video Stabilization, Motion Estimation, HHT, IMFS, SSIM.*

Date of Submission: 20-09-2017 Date of acceptance: 02-10-2017

I. Introduction

Video, being an image sequence, is often called video sequence as well. Video sequences are used within a number of applications such as broadcasting, video-phone, teleconferencing systems, satellite observations or surveillance systems, autonomous navigation, motion analysis, object tracking, astronomical and medical imaging. In the last decade multimedia devices like camcorders, mobile phones have been dramatically increased. Moreover the increasing of their computational performances combined with higher storage capability permits them to store the video formed by the sequence of images. These devices which are typically small and thin have video acquisition capability. However making a stable video with these devices is a very challenging task especially when a zoom lens or a digital zoom is used. Thus, Video 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. 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).

A variety of digital camcorders, 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). The objective of this paper is to eliminate the possible fluctuations due to the jitter motion and obtain the stabilized image sequence with smooth transitions. The Block based Full search motion estimation method is used at the first stage to obtain the local motion vectors for smaller frame regions. After finding local motion vectors, Empirical mode decomposition process is done to obtain the intrinsic mode functions. Hilbert transform is applied to each intrinsic mode function to obtain their energy components. Based on the basic features of the unwanted shaking phenomena (high frequencies and small power contents), intentional and jitter motions are determined, and thus motion compensation is applied in order to eliminate possible fluctuations and produce an image sequence with smoother transitions.

Rest of the paper is organized as follows: section II gives the literature survey. Section III describes the details of proposed approach. Section IV discusses the details of simulation results and section V concludes the paper.

II. Literature Survey

Even before Garret W. Brown applied for a SteadiCam patent in 1977, there were several considerations about compensation for a disturbing blur due to handshake, occurring while holding the camera in the hand. Ever since cameras were so small, that one could take photographs holding them in the hand, tripods and monopods had been utilized to capture sharp images. The development of integrated image stabilizers started in 1980's. Canon was the first manufacturer to introduce an interchangeable zoom lens for 35mm SLR, featuring image stabilization in 1995. Many manufacturers also set off engineering researches and invented their own concepts for stabilization, like CCD-Shift or Digital Stabilizers.

Independent component analysis (ICA) [5] is a computational method for separating a multivariate signal into additive subcomponents by assuming that the subcomponents are non-Gaussian signals and that they are all statistically independent from each other. ICA is a special case of blind source separation. ICA finds the independent components by maximizing the statistical independence of the estimated components. We may choose one of many ways to define independence and this choice governs the form of the ICA algorithm. In this method [6], a novel approach for estimating the global motion between frames using a Curve Warping technique known as Dynamic Time Warping is used. This algorithm guarantees robustness also in presence of sharp illumination changes and moving objects. Digital video stabilization enables to acquire video sequences without disturbing jerkiness by compensating unwanted camera movements. In this method [8], a novel fast image registration algorithm based on block matching is used. Unreliable motion vectors (i.e., not related with jitter movements) are properly filtered out by making use of ad-hoc rules taking into account local similarity, local activity and matching effectiveness. Moreover, a temporal analysis of the relative error computed at each frame has been performed. Reliable information is then used to retrieve inter-frame transformation parameters.

III. Methodology

The implemented Digital image stabilization system consists of three stages:

1. LMV estimation: The block based motion estimation method is used to define the local motion vectors of an image sequence.
2. EMD process for LMVs: The empirical mode decomposition is used to decompose each LMV into finite number of Intrinsic Mode Functions (IMF).
3. Determining the jitter motion: Hilbert transform is applied to each IMF in order to define the energy content. Based on the basic features the last IMF with lower energy content is designated as jitter.

A. LMV Estimation

Block-matching motion estimation is widely used in video coding. The various matching criteria for block-based motion estimation are: 1. Sum of absolute difference (SAD). 2. Mean square error (MSE) and 3.Matching pixel count. For LMV estimation, first a specific image region as shown in the fig 1 is determined from every frame of the image sequence. A square image region of N*N pixels is taken and the intensity value of the pixel at the coordinate (n1,n2) in the frame 'k' current frame is denoted as s(n1,n2,k) where (0≤ n1,n2≤ N-1).

In this method, the matching criteria used is Sum of absolute difference (SAD) and its value of a candidate block of k-1 frame at a displacement (i,j) in the reference frame is given by

$$SAD(i,j) = \frac{1}{N^2}\sum_{n1=0}^{N-1}\sum_{n2=0}^{N-1}[s(n_1,n_2,k) - s(n_1+i,n_2+j,k-1)] \quad (1)$$

The SAD value is computed for every displacement

position (i,j) within a specified search range in the reference image. Thus the displacement which gives minimum SAD value is known as motion vector and is given by

$$[d1, d2] = \arg\min [SAD(i, j)] \quad (2)$$

Where arg min is a function that results in position where the SAD(i,j) is minimized.

After determining the matching criteria, the search motion estimation algorithm must be used. The common search methods for motion estimation are: 1. Full Search (FS) method, 2.Three-Step Search (TSS) method, 3.Four-Step Search (FSS) method, 4.Diamond Search (DS) method. Full search method tests all candidate blocks to determine the best matching block but at the rate of high computational complexity. So TSS, FSS, DS were defined. These methods reduce the number of candidate blocks depending on either specific patterns or the direction of best matching block at each search step but at lower accuracy compared to FS. So for accuracy reasons, FS algorithm is used for the proposed work.

B. EMD process for LMVs

EMD process separates non-stationary data into locally non-overlapping time-scale components. In this method, the original signal is decomposed into a number of orthogonal components called as Intrinsic Mode Functions (IMFs). An IMF function should satisfy the following two conditions:

1. The number of extrema and number of zero crossings must either equal or differ by at most one in the data sets.
2. The mean value of the envelope defined by local maxima and envelope defined by local minima is zero at every point.

In the EMD process, first the local minima and local maxima are identified to form lower and upper envelopes by using cubic spline curve fitting process. Then the mean of the upper and lower envelope is calculated as:

$$m_1(t)=[U(t)+L(t)]/2 \quad (3)$$

Where U(t) and L(t) are local maxima and local minima.

The difference between the original signal x(t) and the mean m_1 is the first component h_1

$$h_1(t)=x(t)-m_1(t) \quad (4)$$

Thus the first component obtained is treated as IMF but there is some error due to the cubic curve spline fitting process. So the shifting process is repeated in order to eliminate the riding waves and to obtain the signal in symmetric form.

In this shifting process, h_1 is treated as data and with the new mean m_{11}, the IMF is computed as

$$h_{11}(t)= h_1(t)-m_{11}(t) \quad (5)$$

After repeating this shifting process up to k times, h_{1k} is computed as

$$h_{1k}(t)= h_{1(k-1)}(t)-m_{1k}(t) \quad (6)$$

Thus h_{1k} is the first IMF which contains the shortest period component of data.

Let $h_{1k}=c_1$ and then $c_1(t)$ is removed from the original signal x(t) to obtain the residual r(t) which contains the information of longer periodic components and is given as

$$r_1(t)=x(t)-c_1(t) \quad (7)$$

As shown in the fig3 the outer loop shifting process is repeated in order to obtain all the subsequent r_w functions as

$$r_w(t)=r_{w-1}(t)-c_w(t) \quad (8)$$

Where w=2,3,……..,n.

The termination criterion for the shifting process is done by considering the SAD value as 0.2. Thus the entire EMD process is terminated if the residue r_w is a monotonic function from which no IMF is obtained.

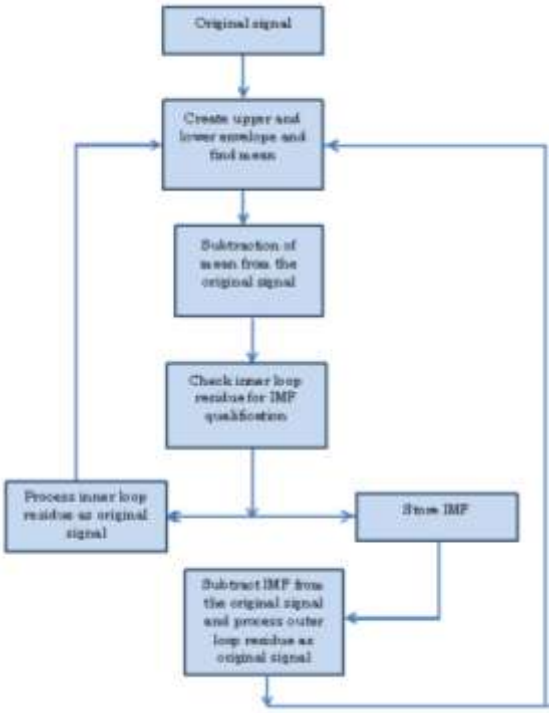

Figure.1. Flow chart of EMD

Thus the EMD process produces a finite number of IMFs and the IMFs with lower indices corresponding to high frequency displacements. But to obtain the IMF which contains jitter their energy component should be determined.

C. Determining the jitter motion

The intentional motion of the camera contains smooth transitions and this motion has higher pixel displacements leading to higher energy levels compared to jitter motion. The jitter motion contains lower pixel displacements leading to lower energy levels. Since the EMD process divides the original signal into number of sub signals based on their frequencies, the last IMF which includes jitter should be determined. So Hilbert transform is applied to each IMF to obtain the energy contents. Thus after applying the Hilbert transform the instantaneous amplitude and the instantaneous frequency of the signal are given by

$$\alpha(t)=\sqrt{x(t)^2 + y(t)^2} \quad (9)$$

$$\theta(t)=\tan^{-1}\left(\frac{y(t)}{x(t)}\right) \quad (10)$$

where x(t) is the original signal and y(t) is the Hilbert transform of x(t).

The power of each IMF is proportional to the amplitude and is given as

$$P_i=\sum_{t=0}^{k}(\alpha_i(t))^2 \qquad (11)$$

Where α_i denotes the amplitude of IMFs sample, i=1,2,……..,w+1 denotes the corresponding IMF and t denotes the number of frame.

Thus the IMF with the higher index and the lower energy content is determined as the last IMF which includes jitter components and the required threshold is given as

$$d= \arg\min[P_i] \qquad (12)$$

where i=1,2,…..,w+1 and arg min represents the position where P_i is minimized.

Thus the summation of all the IMFs up to the threshold d defines the unwanted camera motion and the jitter ($X_j(t)$)is given by

$$X_j(t) = \sum_{i=1}^{d}(c_i(t)) \qquad (13)$$

The summation of the remaining IMFs including the residue gives the intentional motion of the camera ($X_g(t)$) and is given by

$$X_g(t)= \sum_{i=d}^{w}(c_i(t))+ r_w \qquad (14)$$

IV. Simulation Results

To verify the effectiveness of the implemented DIS method, several simulations were performed, and the results were compared with existing stabilization method. As it was mentioned above, only vertical displacements are presented since the procedure for horizontal motions is exactly the same. In order to evaluate the performance of the method, two different image sequences were processed. The most widely used metric for such applications is the Root Mean Square Error (RMSE) which was applied so that a quantitative comparison with the other related methods could be achieved and is calculated by

$$e_{rms} = \frac{1}{N}\sqrt{\sum_{n=1}^{N}(\bar{x}_n - x_n)^2 + (\bar{y}_n - y_n)^2} \qquad (15)$$

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. 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. For the presented experiments, a servo mechanism along with an attached handheld camera was used in order to produce image sequences with known motions.

The servo was rotated according to the wanted (intentional and jitter) camera motion while the camera was capturing the scene frames. Since the camera motion is reflected in the acquired image sequence, its motion could be retrieved using the calculated LMVs which include both the intentional and the jitter motions. The size of the squared region was randomly selected to be 16 X 16 pixels. In order to produce LMVs, the Full-search Block matching algorithm was used for achieving the most accurate resulting signal, despite the increased time complexity. The estimated LMV is decomposed into a finite number of IMFs by applying the EMD process. As terminating criterion of the sifting process, an SD value equal to 0.2 was used according to Huang's analysis. Hilbert transform is then applied to each sub signal in order to compute their energy content. Depending on the resulting values, the last IMF to be considered as a noise sub signal is defined. The summation of the prime IMFs up to the defined threshold value of 9 represents the jitter motion, and thus, the remaining signal signifies the intentional motion of the camera.

Figure.2 Rhino sample

The simulation results for Rhino sample are shown in below figures. The Rhino sample contains total of 114 frames and each frame at a rate of 15 frames/sec. This sample has a frame size of 240×320 and the duration of video sample is 7sec. the first four uncompensated frames of Rhino sample containing jitter motion are shown in below figures.

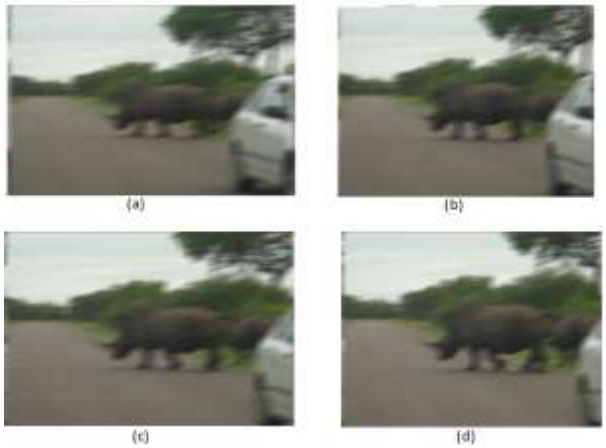

Figure.3:(a),(b),(c)&(d) Consecutive four frames of Rhino sample

After applying Hilbert transform to each IMF, their energy content is calculated and the last IMF containing jitter is determined based on the threshold value equal to 9 which contains minimum power value. Thus the summation of IMFs up to this value is eliminated using motion compensation and the summation of all IMFs contains intentional camera motion. The first four compensated frames are shown in below figures.

Figure.4:(a)1st Jitter frame (b) 1st Compensated frame

Figure.5:(a)2nd Jitter frame (b) 2nd Compensated frame

Figure.6:(a)3rd Jitter frame (b) 3rd Compensated frame

Figure.7:(a)4th Jitter frame (b) 4th Compensated frame

The RMSE values for each frame are shown in tabular column:

Table.1: RMSE values of first four frames of Rhino sample

Frame number	RMSE value
Frame 1	0.0784
Frame 2	0.0786
Frame 3	0.0785
Frame 4	0.0785

The simulation results for Water drop sample are shown in below figures. The Water drop sample contains total of 182 frames and each frame at a rate of 30 frames/sec. This sample has a frame size of 240×256 and the duration of video sample is 6sec. The first four uncompensated frames of Water drop sample containing jitter motion are shown in below figures.

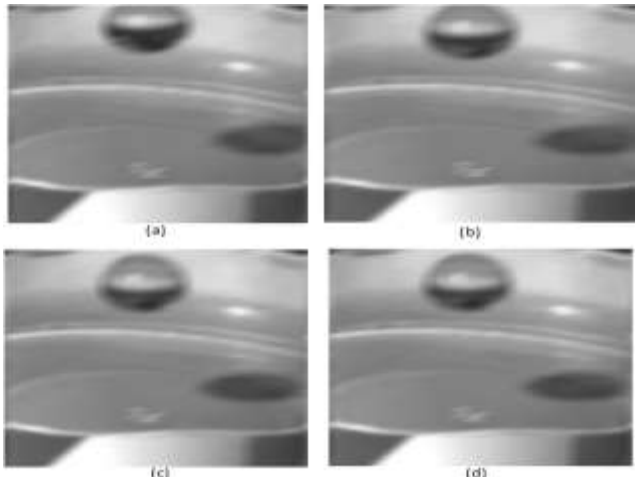

Figure 8:(a),(b),(c)&(d) Consecutive four frames of water drop sample

After applying Hilbert transform to each IMF, their energy content is calculated and the last IMF containing jitter is determined based on the threshold value equal to 9 which contains minimum power value. Thus the summation of IMFs up to this value is eliminated using motion compensation and the summation of all IMFs contains intentional camera motion. The first four compensated frames are shown in below figures.

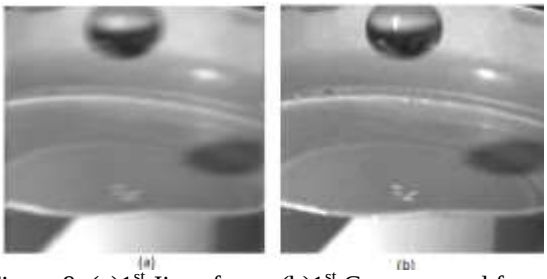

Figure 9: (a)1st Jitter frame (b)1st Compensated frame

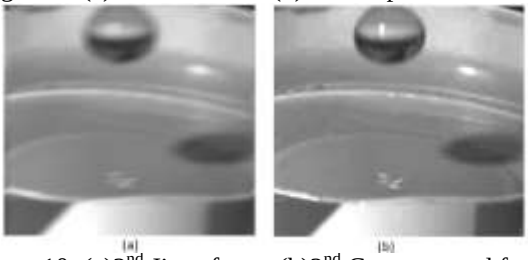

Figure 10: (a)2nd Jitter frame (b)2nd Compensated frame

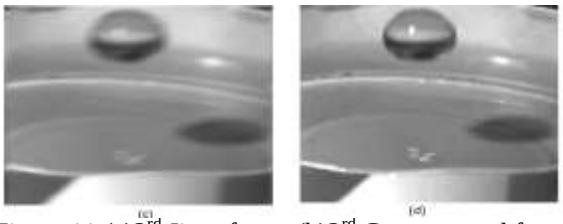

Figure 11:(a)3rd Jitter frame (b)3rd Compensated frame

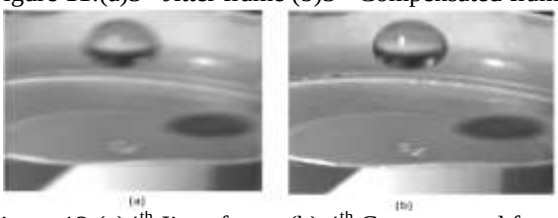

Figure 12:(a)4th Jitter frame (b) 4th Compensated frame

The RMSE values for each frame are shown in following tabular column:

Table.2: RMSE values of first four frames of water drop sample

Frame number	RMSE value
Frame 1	0.0341
Frame 2	0.0343
Frame 3	0.0342
Frame 4	0.0341

In the fig.8, the displacement is high at frame 22. Therefore, the jitter and compensated 22nd frame is shown in below figure:

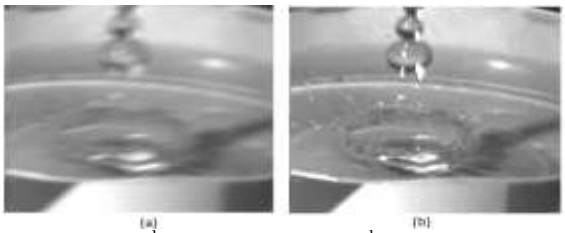

Figure 13:(a)22nd Jitter frame (b) 22nd Compensated frame

The compensated video samples and its RMSE values using existing DIS method and implementing DIS method by HHT are shown in below figures.

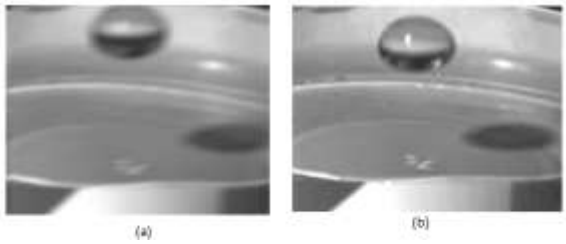

Figure 14: Compensated water drop sample by (a) DIS (b) HHT

The comparison of RMSE values for DIS and DIS by HHT method for two input samples is shown in below tabular column:

Table.3: RMSE values of two DIS methods

Method	**S1**	**S2**
Conventional	0.31312	0.13577
Proposed	0.078293	0.033959

V. Conclusion

In this work, the Digital image Stabilization using Hilbert-Huang transform, different video samples containing colour and black and white samples captured by the digital camcorder are processed to produce the image sequence with smooth transitions. The effectiveness of the method relies on the fact that the jitter signal is

defined based on its two principal features, high frequencies and low-energy content. The complete implementation of this project is done using Matlab 2012. The quantitative performance criteria used to compare with the existing DIS is Root Mean Square Error (RMSE). Simulation results have shown that the implemented method successfully decomposes the two camera motions, and the image sequence is compensated effectively. The results also shows that the implemented method exhibits better performance based on the lower RMSE values when compared with the existing system. In this project, motion estimation using Block-based Full search method is used. This Full search method uses sum of absolute difference and mean square error as matching criteria which takes more time but have more accuracy. So, in the future, we may develop new motion estimation method which reduces the computational time.

References

[1]. C. Caraffi, S. Cattani, and P. Grisleri, "Off-road path and obstacle detection using decision networks and stereo vision," IEEE Trans. Intell. Transp. Syst., vol. 8, no. 4, pp. 607–618, Dec. 2007.

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

[3]. A. Bosco, A. Bruna, S. Battiato, G. Bella, and G. Puglisi, "Digital video stabilization through curve warping techniques," *IEEE Trans. Consum. Electron.*, vol. 54, no. 2, pp. 220–224, May 2008.

[4]. A.-O. Boudraa and J.C. Cexus, "EMD-based signal filtering," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2196–2202, Dec. 2007.

[5]. S. Battiato, A. R. Bruna, and G. Puglisi, "A robust block based image/video registration approach for mobile imaging devices," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 622–635, Nov. 2010.

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

[7]. B. E. Burke, R. K. Reich, E. D. Savoye, and J. L. Tonry, "An orthogonal-transfer CCD imager," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2482–2484, Dec. 1994.

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

[9]. R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," IEEE Trans. Circuits Syst. Video Technol., vol. 4, no. 4, pp. 438–442, Aug. 1994.

[10]. L. M. Po and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation," IEEE Trans. Circuits Syst. Video Technol., vol. 6,no. 3, pp. 313–317, Jun. 1996.

[11]. S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block matching motion estimation," IEEE Trans. Image Process., vol. 9, no. 2,pp. 287–290, Feb. 2000.

[12]. 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.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

B. Sowbhagya . "An Enhanced Video Stabilization Based On Emd Filtering And Spectral Analysis ." IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 5, 2017, pp. 23–30.