Trans. Tianjin Univ. 2011, 17: 068-074
DOI 10.1007/s12209-011-1474-y

Distortion Identification Technique Based on Hilbert-Huang Transform in Video Stabilization*

LIU Yan(刘　艳)[1], ZOU Mouyan(邹谋炎)[1,2], WANG Qiang(王　强)[3]
(1. Graduate University, Chinese Academy of Sciences, Beijing 100190, China;
2. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China;
3. School of Sciences, Tianjin University, Tianjin 300072, China)

Abstract: A distortion identification technique is presented based on Hilbert-Huang transform to identify distortion model and distortion frequency of distorted real-world image sequences. The distortion model is identified simply based on Hilbert marginal spectral analysis after empirical mode decomposing. And distortion frequency is identified by analyzing the occurrence frequency of instantaneous frequency components of every intrinsic mode functions. Rational digital frequency filter with suitable cutoff frequency is designed to remove undesired fluctuations based on identification results. Experimental results show that this technique can identify distortion model and distortion frequency of displacement sequence accurately and efficiently. Based on identification results, distorted image sequence can be stabilized effectively.

Keywords: image sequence distortion; video stabilization; distortion model identification; distortion frequency identification; Hilbert-Huang transform

Video shot of a digital camera usually suffers from various distortions due to unstable random camera motion. Video stabilization technique removes undesired fluctuations, estimates intentional global motions of the camera and compensates them so that we can remove or reduce those jitter fluctuations' influence. Dynamic displacement field model is a parameterized model and provides a gradually upgraded and evolutionary process, from simple to complex, to handle image sequence distortion problems[1]. Since jitter distortion is a kind of high-frequency variation of image displacements, conventional frequency domain filter can be used to remove undesired fluctuations[1,2]. According to the relationship between distortion frequency and motion model, video stabilization becomes a facile parameter estimation process. And it is crucial to identify distortion model and distortion frequency of the acquired displacement sequence before filtering. Based on identification results, motion filtering and motion compensation can be used to suppress jitter[1-6]. However, displacement sequence is usually non-stationary because of unstable random camera motion. Conventional Fourier analysis is suitable for linear stationary signal, while it is not suitable for nonlinear non-stationary signal.

This paper proposes a distortion identification technique based on Hilbert-Huang transform (HHT)[7,8] to identify distortion model and distortion frequency of real-world image sequence. Firstly, displacement sequence is decomposed into a set of intrinsic mode functions by empirical mode decomposition method. Secondly, distortion model and distortion frequency are identified based on Hilbert spectral analysis. Finally, a rational digital filter with suitable cutoff frequency is designed based on identification results. Based on these identification results, undesired fluctuation can be removed with the stabilization methods presented in Refs.[1] and [2].

1 Dynamic displacement field model

Dynamic displacement field (DDF) model describes the dynamic displacement of each pixel of an image relating to its original reference. The mathematical description of DDF model is[1]

$$B(i', j'; t) = A(i', j'; t) + \xi(i', j'; t) \tag{1}$$

with

$$i' = i + \sum_{k=1}^{K} x_k(i, j; t)$$

Accepted date: 2010-06-25.
*Supported by the President Fund of Graduate University, Chinese Academy of Sciences.
LIU Yan, born in 1979, female, Dr, lecturer.
Correspondence to LIU Yan, E-mail: yanliu@gucas.ac.cn.

$$j' = j + \sum_{k=1}^{K} y_k(i,j;t)$$
$$i \in [1,M];\ j \in [1,N] \qquad (2)$$

where A is the original image whose size is $M \times N$; B the new image by displacing each pixel of A from location (i,j) to new location (i',j') with an additive noise $\xi(i',j';t)$; $x_k(i,j;t)$ and $y_k(i,j;t)$ are spatial displacements of the pixel $A(i,j;t)$ at time t in i and j direction, respectively. A pair $(x_k(i,j;t), y_k(i,j;t))$ expresses a displacement pattern. In Eq.(2), we suppose that there are K different displacement patterns. The variable t stands for continuous or discrete time. Eq.(2) implies that displacement $x_k(i,j;t)$ and $y_k(i,j;t)$ are time sequences and they may be spatially variable.

Eq.(2) provides a decomposable description for complex and dynamic distortions. The subscript k can be changed to have a specified meaning. For example, $(x_{\text{jit}}(i,j;t), y_{\text{jit}}(i,j;t))$ can be used to describe jittering displacements. At each moment, if the jittering displacements of all pixels are (statistically) equal, jittering is (statistically) spatially invariable and the pair $(x_{\text{jit}}(i,j;t), y_{\text{jit}}(i,j;t))$ can be simplified to $(x_{\text{jit}}(t), y_{\text{jit}}(t))$. Similarly, $(x_{\text{sh}}(t), y_{\text{sh}}(t))$ is used to express the spatially invariable shift and such a distortion model is called a zero-order model or translation model. In most cases, the spectra of translation displacements are within a low-frequency band. Contrarily, the spectra of jittering displacements are within a high-frequency band.

Based on the parameterized DDF model, video stabilization becomes a facile parameter estimation process. According to the relationship between distortion frequency and motion model, conventional filtering methods will be used to suppress jitter of the estimated displacements[1,2].

2 Identification technique based on HHT method

To make up the deficiency of Fourier analysis, time-frequency analysis is proposed for non-stationary signal[7-10]. In 1998, Huang *et al* proposed an empirical mode decomposition (EMD) method for adaptively decomposing signal[7]. EMD method decomposes signal into a set of functions defined by itself, i.e., the intrinsic mode functions (IMFs). Based on EMD method, HHT method was given as a new time-frequency analysis. HHT makes the instantaneous frequency meaningful by EMD method and presents the result of time-frequency analysis in a Hilbert spectrum plot[7,8]. Considering the validity and rationality of HHT method used for analyzing non-stationary signal[11-15], this paper presents an identification technique based on HHT method to identify distortion model and distortion frequency, but discuss little about HHT method itself.

2.1 HHT method and Hilbert spectral analysis

Assuming that EMD method expands signal $x(t)$ into n IMFs with residue r_n, we obtain a decomposed signal $x(t) = \sum_{k=1}^{n} c_k + r_n$. After making Hilbert transform for every IMF and accumulating them, $x(t)$ can be expressed as

$$x(t) \approx \mathrm{Re}\sum_{k=1}^{n} a_k(t)\,\mathrm{e}^{\mathrm{j}\theta_k(t)} = \mathrm{Re}\sum_{k=1}^{n} a_k(t)\,\mathrm{e}^{\mathrm{j}\int \omega_k(t)\mathrm{d}t} \qquad (3)$$

where $a_k(t)$ and $\theta_k(t)$ are amplitude and phase of the kth IMF component, respectively; $\omega_k(t)$ is its corresponding instantaneous frequency.

Hilbert spectrum $H(\omega,t)$ can be defined as Eq.(4), which shows the local information of time, frequency and energy.

$$H(\omega,t) = \mathrm{Re}\sum_{k=1}^{n} a_k(t)\,\mathrm{e}^{\mathrm{j}\int \omega_k(t)\mathrm{d}t} \qquad (4)$$

Hilbert marginal spectrum represents the contribution of every frequency component[7,8], which is defined as

$$h(\omega) = \int H(\omega,t)\,\mathrm{d}t \qquad (5)$$

where the higher value of $h(\omega)$ reflects the greater occurrence frequency of ω .

2.2 Identification technique based on HHT method

HHT method uses EMD method to decompose non-stationary signal into a set of stationary IMFs and analyzes every IMF by Hilbert spectral analysis to obtain instantaneous frequency of every IMF[7,8]. Based on the instantaneous frequency and instantaneous amplitude, the dynamic characteristics of distortion frequency are analyzed.

EMD method always extracts the primary information of signal firstly, i.e., several initial IMFs including the main information of signal. Therefore, according to the contribution of every frequency component in the whole spectrum, the main distortion characteristic of displacement sequence can be identified using Hilbert marginal spectral analysis.

Step 1 Decomposing displacement sequence to obtain n IMFs using EMD method, which are represented as $c_k(k = 1,\cdots,n)$.

Step 2 Analyzing every IMF based on Hilbert spectral analysis to obtain Hilbert spectrum $H_k(\omega,t)$ and instantaneous frequency $\omega_k(t)$.

Step 3 Dividing the spectrum $[\min(\omega(t)), \max(\omega(t))]$ of instantaneous frequency $\omega(t)$ into N_1 segments with interval $\Delta\omega = [\max(\omega(t)) - \min(\omega(t))]/N_1$.

Step 4 Accumulating instantaneous amplitude whose corresponding instantaneous frequency belongs to a specified spectral segment. After processing every segment like this, Hilbert marginal spectrum $h(\omega)$ of signal is obtained.

Step 5 Analyzing marginal spectrum $h(\omega)$. The energy concentrated frequency range reflects the main distortion characteristic of signal. If energy mainly exists in high-frequency band, there are serious jitter distortions in image sequence; accordingly, there are slow low-frequency translations in sequence if energy is mainly located in low-frequency band. Thus, distortion model is identified simply.

To reduce or remove undesired fluctuations, it is necessary to design a digital filter based on identified distortion frequency. There is a marginal spectrum analysis which uses instantaneous frequency corresponding to the maximum amplitude as the cutoff frequency of filter. But it will bring some estimation error because the marginal spectrum often has several peak values. Since distortion frequency identification aims to find the primary frequency component of displacement sequence, distortion frequency will be identified by analyzing the occurrence frequency of instantaneous frequency components of every IMF. Based on HHT method, this paper proposes a simple distortion frequency identification method as follows, which can reduce estimation error and identify distortion frequency correctly and easily. Assume that a set of estimated IMF components $c_k(k=1,\cdots,n)$ of displacement sequence and their corresponding Hilbert spectra $H_k(\omega,t)(k=1,\cdots,n)$ have been obtained.

Step 1 Analyzing instantaneous frequency components $\omega_k(t)$ of every IMF and computing their occurrence frequency.

Step 2 Choosing frequency components $f_k(k=1,\cdots,n)$ of every IMF, which has the biggest occurrence frequency among other frequency components.

Step 3 Adding all chosen frequency components together and averaging the sum to obtain the estimated distortion frequency $\hat{f} = \frac{1}{n}\sum_{k=1}^{n} f_k$.

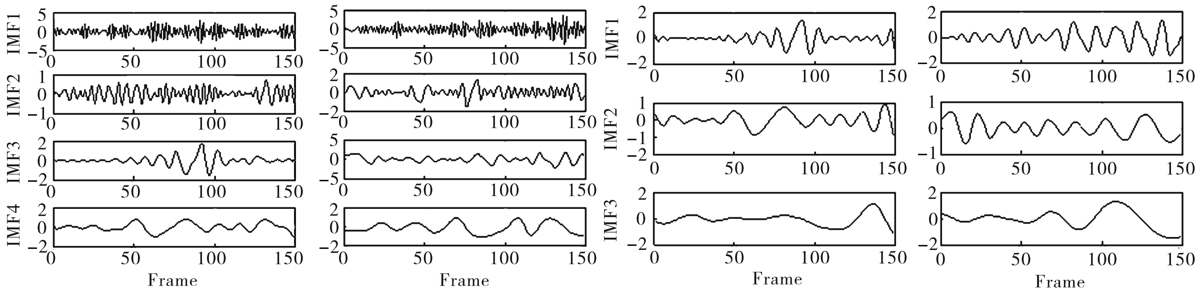
Based on the identified distortion model and distortion frequency, undesired fluctuations can be removed with the stabilization methods presented in Ref.[1] and Ref.[2].

3 Experimental results

Firstly, the simulation results of DDF model in Ref.[1] are taken as an example to explain the efficiency of HHT method for identifying distortion model and distortion frequency. In DDF model, the distortion frequency of jittering model and translation model in x and y direction is initialized as

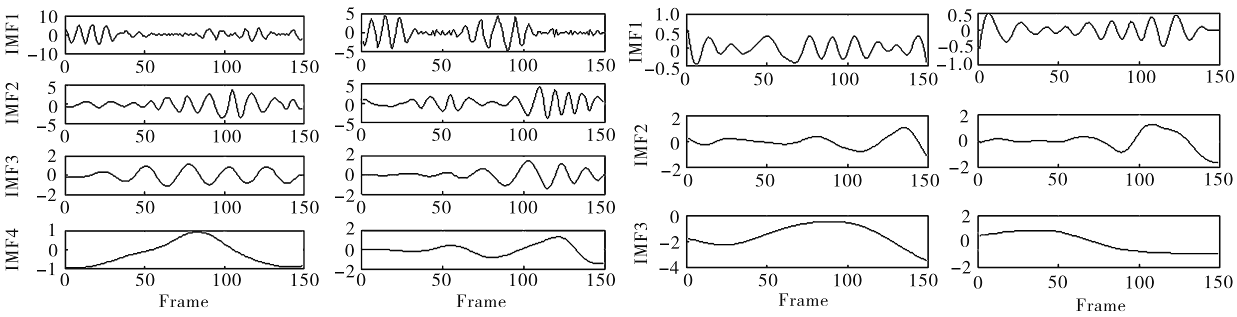
$$f_{jx} = f_{jy} = 6\text{ Hz},\ f_{shx} = f_{shy} = 2\text{ Hz}.$$

Assuming we have acquired jittering displacement sequences $(x_{jit}(t), y_{jit}(t))$ and translation displacement sequences $(x_{sh}(t), y_{sh}(t))$. Next, decompose four displacement sequences using EMD method to obtain IMFs shown in Fig.1(a),(b) and Fig.2(a),(b). Then, Hilbert marginal spectra of these four sequences are shown in Fig.3. Fig.3(a) and Fig.3(b) show that energy is mainly located in high-frequency band, which explains why the estimated sequence contains high-frequency jitter and is consistent with the characteristic of jittering displacement field model. Similarly, in Fig.3(c) and Fig.3(d), the estimated sequence contains low-frequency translation, which is consistent with the characteristic of zero-order displacement field model. The identification results are consistent with the simulation settings.



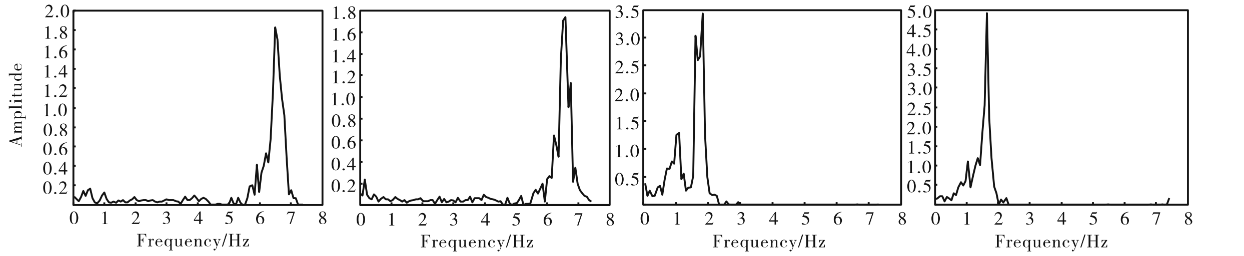
(a) IMFs in horizontal direction before stabilization (b) IMFs in vertical direction before stabilization (c) IMFs in horizontal direction after stabilization (d) IMFs in vertical direction after stabilization

Fig.1 EMD decomposition results of the estimated jittering displacement sequence of a car



(a) IMFs in horizontal direction before stabilization (b) IMFs in vertical direction before stabilization (c) IMFs in horizontal direction after stabilization (d) IMFs in vertical direction after stabilization

Fig.2 EMD decomposition results of the estimated translation displacement sequence of a car



(a) Jittering displacement sequence in horizontal direction (b) Jittering displacement sequence in vertical direction (c) Translation displacement sequence in horizontal direction (d) Translation displacement sequence in vertical direction

Fig.3 Hilbert marginal spectra of jittering displacement and translation displacement sequences

After computing occurrence frequency of instantaneous frequency components of every IMF, the cutoff frequency of motion filter can be set based on the estimated frequency $\hat{f}$. During motion filtering, frequency components higher than $\hat{f}$ will be filtered. Compared with the results identified by Hilbert marginal spectrum, the identification results using the proposed method are closer to the setting frequency (see Tab.1). From Fig.1(c), (d) and Fig.2(c), (d), high-frequency jitter is effectively suppressed.

Tab.1 Identification results of distortion frequency of DDF model

Experimental sequence	Distortion frequency identified by Hilbert marginal spectrum /Hz		Distortion frequency identified by proposed method /Hz	
	x direction	y direction	x direction	y direction
Jittering displacement sequence	6.494 1	6.598 1	6.196 1	6.360 5
Translation displacement sequence	1.830 1	1.649 9	2.133 6	1.813 5

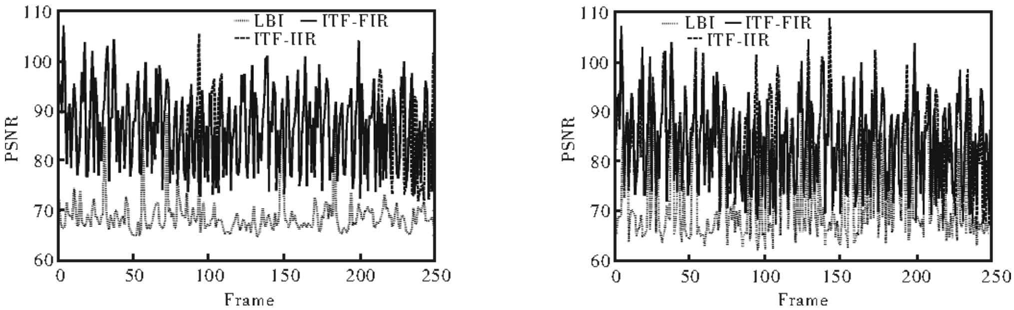
In this paper, the interframe transformation fidelity (ITF) is used to reflect the performance of stabilization, which is defined by peak signal to noise ratio (PSNR) of successive frames in stabilized sequences. The lower bound value of ITF is defined as LBI. LBI and ITF are defined as[16]

$$\mathrm{LBI} = \mathrm{PSNR}(I_k, I_{k-1}) = 10\ \mathrm{lg}\frac{255^2}{\mathrm{MSE}(I_k, I_{k-1})} \tag{6}$$

$$\mathrm{ITF} = \mathrm{PSNR}(S_k, S_{k-1}) = 10\ \mathrm{lg}\frac{255^2}{\mathrm{MSE}(S_k, S_{k-1})} \tag{7}$$

where I_k is the kth frame of sequence before stabilization; S_k the kth frame of sequence after stabilization; MSE (I_k, I_{k-1}) the mean square error between I_{k-1} and I_k; MSE (S_k, S_{k-1}) the mean square error between S_{k-1} and S_k.

By comparing the values of LBI and ITF of image sequence, the performance of stabilization can be evaluated. For example, the value of ITF will be larger than LBI for stabilized sequence, while unsuccessful stabilization will produce lower ITF value. Some experimental results based on jittering model and translation model of a car are shown in Fig.4, where ITF-FIR and ITF-IIR show the value of ITF when using conventional FIR filter or IIR filter in stabilization process described in Ref.[1] and Ref.[2]. It can be seen that the values of ITF are higher than those of LBI, showing that the proposed stabilization techniques are more effective in suppressing higher-frequency jitter and removing lower-frequency translation, also proving that the proposed identification technique can identify distortion model and distortion frequency accurately and effectively.

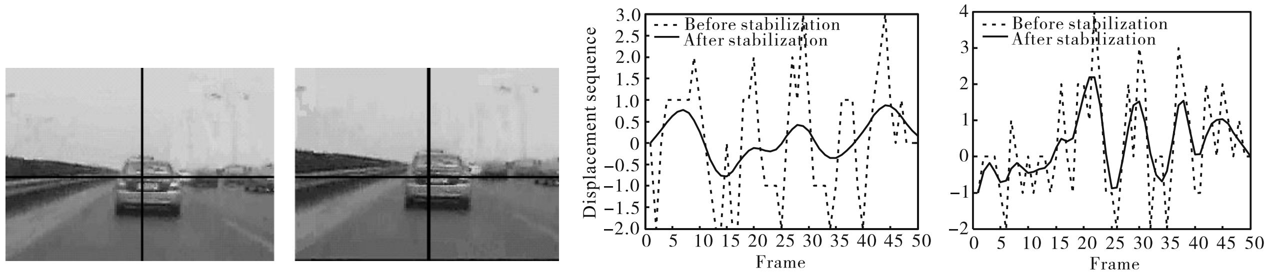


(a) For jittering model (b) For translation model

Fig.4 PSNR curve for jittering model and translation model of the car

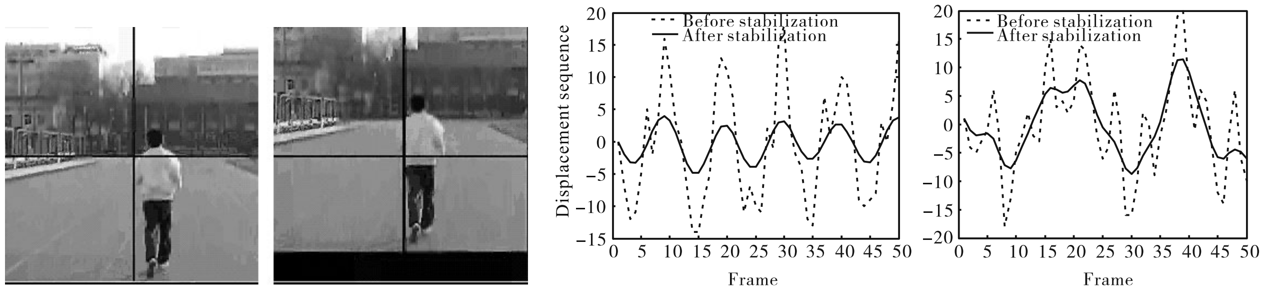
Based on the efficiency of HHT method in distortion frequency identification, the proposed method is used to identify distortion frequency of real-world video sequences (Fig.5(a) and Fig.6(a)). From the displacement curves before and after stabilization (Fig.5(c), (d) and Fig.6(c), (d)), high-frequency jitter in vertical direction is preferably suppressed. From the PSNR curves before and after stabilization (Fig.7), the values of ITF for both sequences are higher than those of LBI. From several typical frames of experimental results (Fig.8(b), (d) and Fig.9(b), (d)), the objects (car in Fig.8 and person in Fig.9) are located in the middle of the image after stabilization. A few seconds of the stabilization results in AVI format are available in Ref.[17] and Ref.[18].

The stabilization result of real-world sequence 2 using commercial video stabilization software Steadyhand is also included in Ref.[17] and Ref.[18]. In video sequences, jitter distortion caused by unstable camera motion is eliminated and our experimental results are more steady than results stabilized by Steadyhand software.



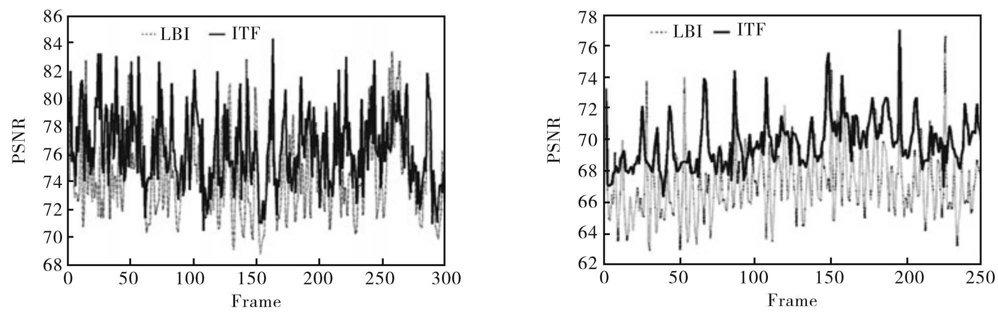
(a) Original image sequence (b) Stabilized image sequence (c) Displacement sequence in horizontal direction (d) Displacement sequence in vertical direction

Fig.5 Stabilization results of real-world sequence 1



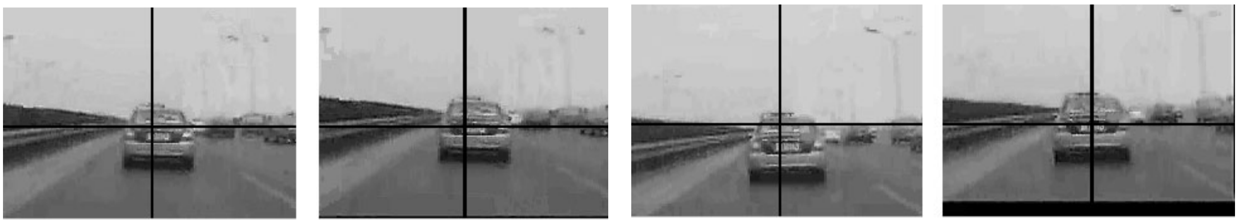
(a) Original image sequence (b) Stabilized image sequence (c) Displacement sequence in horizontal direction (d) Displacement sequence in vertical direction

Fig.6 Stabilization results of real-world sequence 2



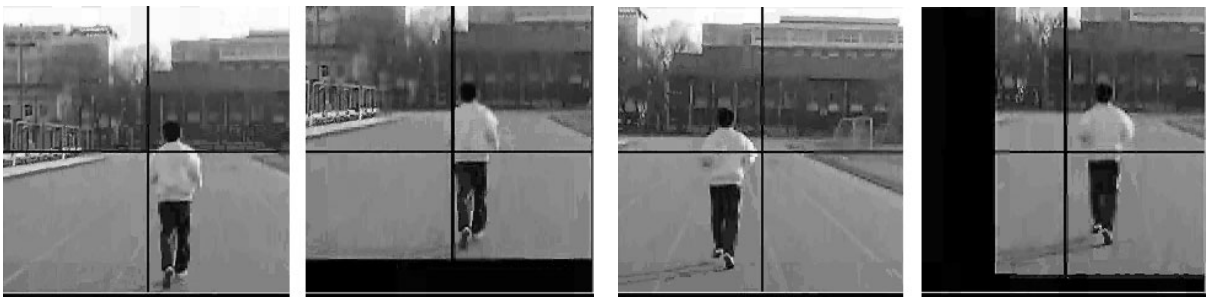
(a) For real-world sequence 1 (b) For real-world sequence 2

Fig.7 PSNR curves for real-world sequence 1 and sequence 2



(a) The 70th frame before stabilization (b) The 70th frame after stabilization (c) The 150th frame before stabilization (d) The 150th frame after stabilization

Fig.8 Typical frames of real-world sequence 1 before and after stabilization



(a) The 50th frame before stabilization (b) The 50th frame after stabilization (c) The 150th frame before stabilization (d) The 150th frame after stabilization

Fig.9 Typical frames of real-world sequence 2 before and after stabilization

The identification results before and after stabilization are shown in Tab.2. Most jittering frequencies are reduced after stabilization (see Tab.2). It is shown that motion filter whose cutoff frequency set by identified frequency based on HHT method can efficiently suppress jittering distortion of image sequence.

Tab.2 Identification results of distortion frequency before and after stabilization

Experimental sequence	Distortion frequency before stabilization/Hz		Distortion frequency after stabilization/Hz	
	x	*y*	*x*	*y*
Jittering model	6.196 1	6.360 5	4.188 6	4.274 5
Car sequence	1.300 6	3.294 3	1.247 6	1.695 1
Walk sequence	1.602 4	2.341 9	1.334 3	1.056 6

4 Conclusions

This paper proposes an identification technique based on HHT method to identify distortion model and distortion frequency of sequence. Distortion model was identified simply by identifying the main distortion characteristic of displacement sequence using Hilbert marginal spectral analysis. Distortion frequency was identified by analyzing the occurrence frequency of instantaneous frequency components of every IMF, which reduced estimation error and identified distortion frequency correctly and easily. Based on identification results, motion filter used for video stabilization was designed and the performance of stabilization was verified by ITF. Experimental results show that this technique identifies distortion model and distortion frequency of real-world sequence correctly and efficiently. Based on identification results, undesired fluctuations can be removed efficiently.

References

[1] Liu Yan, Zou Mouyan, Cao Ying *et al*. Dynamic displacement field model used as a new camera motion model in video stabilization[C]. In: *Proceedings of International Conference on Consumer Electronics*. Las Vegas, USA, 2007. 1-2.

[2] Liu Yan, Zou Mouyan. Video stabilization technique for spatially variable distorted image sequence based on dynamic displacement field model[J]. *Journal of the Graduate School of the Chinese Academy of Sciences*, 2008, 25(1): 117-122.

[3] Chang Hungchang, Lai Shanghong, Lu Kuangrong. A robust and efficient video stabilization algorithm[C]. In: *Proceedings of IEEE International Conference on Multimedia and Expo*. Taipei, China, 2004. Vol. 1. 29-32.

[4] Zerafat Pisheh M A, Sheikhi A. Detection and compensation of image sequence jitter due to an unstable CCD camera for video tracking of a moving target[C]. In: *Proceedings of the 2nd International Symposium on 3D Data Processing*, *Visualization and Transmission*. Thessaloniki, Greece, 2004. 258-261.

[5] Pan J, Hu B, Zhang H Q. Robust and accurate object tracking under various types of occlusions[J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2008, 18(2): 223-236.

[6] Farsiu S, Elad M, Milanfar P. Multiframe demosaicing and super-resolution of color images[J]. *IEEE Trans on Image Processing*, 2006, 15(1): 141-159.

[7] Huang N E, Shen Z, Long S R *et al*. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society of London Series A Mathematical Physical and Engineering Sciences*, 1998, 454(1971): 903-995.

[8] Huang N E, Wu M-L C, Long S R *et al*. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis[J]. *Proceedings of the Royal Society of London Series A Mathematical Physical and Engineering Sciences*, 2003, 459(2037): 2317-2345.

[9] Stankovic L. A method for time-frequency analysis[J]. *IEEE Trans on Signal Processing*, 1994, 42(1): 225-229.

[10] Hlawatsch F, Boudreaux-Bartels G F. Linear and quadratic time-frequency signal representations[J]. *IEEE Signal Processing Magazine*, 1992, 9(2): 21-67.

[11] Roy A, Chun-Hsien Wen, Doherty J F *et al*. Signal feature extraction from microbarograph observations using the Hilbert-Huang transform[J]. *IEEE Trans on Geoscience and Remote Sensing*, 2008, 46(5): 1442-1447.

[12] Messina A R, Vittal V. Extraction of dynamic patterns from wide-area measurements using empirical orthogonal functions[J]. *IEEE Trans on Power Systems*, 2007, 22(2): 682-692.

[13] Wang Hengliang, Siu Kin, Ju Kihwan *et al*. Identification of transient renal autoregulatory mechanisms using time-frequency spectral techniques[J]. *IEEE Trans on Biomedical Engineering*, 2005, 52(6): 1033-1039.

[14] Flandrin P, Rilling G, Goncalves P. Empirical mode decomposition as a filter bank[J]. *IEEE Signal Processing Letters*, 2004, 11(2): 112-114.

[15] Wu F J, Qu L S. An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery[J]. *Journal of Sound and Vibration*, 2008, 314(3-5): 586-602.

[16] Carlos Morimoto, Rama Chellappa. Evaluation of image stabilization algorithms[C]. In: *Proceedings of IEEE International Conference on Acoustics*, *Speech and Signal Processing*. Seattle, USA , 1998. Vol. 5. 2789-2792.

[17] Liu Yan, Zou Mouyan. Distortion identification technique in video stabilization[C]. In: *Proceedings of International Conference on Consumer Electronics*. Las Vegas, USA, 2009. 1-2.