Digital image stabilization method based on variational mode decomposition and sampling fluctuation analysis

Duo Hao[1,a], Wencong Shi[1,b], Chengwei Li[2,c,*], Fenghuan Fan[1,d]

[1]Department of Electronic Information and Control Engineering, North China Institute of Aerospace Engineering, Hebei, Langfang, China
[2]School of Instrument Science and Engineering, Harbin Institute of Technology, Heilonjiang, Harbin, China
[a] haoduonciae@163.com, [b]13073168831@163.com, [c] lichengweihit@163.com,
[d]13833463175@163.com
*Corresponding author

Abstract: *Unintentional motions often cause cameras to produce shaky images, which is a significant source of inter-frame blur and video quality decline. To ad-dress this issue, we present a digital image stabilization approach based on variational mode decomposition (VMD) and sampling fluctuation analysis (SFA) to generate stable video sequences. Our method first estimates the global motion vector (GMV) from a video sequence using the speeded up robust features (SURF) algorithm. We then decompose the GMV into various modes using VMD to separate jitter motions from intentional ones. Here, SFA is applied to distinguish different modes based on their unique structural characteristics. We evaluate our proposed method in complex scenarios by comparing it with several existing methods. Our experimental results demonstrate that VMD outperforms other stabilization techniques under comparable conditions.*

Keywords: digital image stabilization, variational mode decomposition, sampling fluctuation analysis, global motion vector sequence

1. Introduction

The camera equipment is often disturbed[1][2][3][4][5] by the motion of various factors such as terrain and human, which makes the video sequence contain random jitter. Inter-frame motion blur caused by random jitter will not only reduce the accu-racy of observation, affect the subsequent detection, recognition and tracking steps, but also cause visual fatigue to the operator. A typical DIS system consists of three main stages: motion estimation, motion separation, and motion compensa-tion. In the motion estimation stage, the global motion vector (GMV) is es-timated within a specific frame region, which contains both intentional and jitter motion. The motion separation stage is crucial, wherein a certain algo-rithm is employed to separate the intentional and jitter motion. This stage is the primary research focus of this study. Finally, in the motion compensa-tion stage, frames are adjusted to generate a stable sequence post-processing, as illustrated in Figure 1.

In the field of signal processing, camera jitter induced by external factors can be viewed as a noisy signal overlaid on intentional motion vectors. Consequently, several smoothing-based methods for motion separation have been suggested. Mean filtering is a simple low-pass filtering mathematical model, which is one of the widely used schemes[6]. The utilization of a sliding window to smooth the global motion vector (GMV) enables the proposed method to alleviate fluctuation interference arising from accidental factors to a certain degree, but the fixed sliding window is difficult to adapt to the complex application environment of electronic image stabili-zation. Wavelet decomposition (WD) is also a good filtering method, which is a time-frequency domain filtering technique proposed to meet nonlinear condi-tions[7][8]. The WD method requires the selection of an appropriate wavelet basis function beforehand, which can be challenging in complex application environments. Recently, a motion estimation technique based on empirical mode decomposition (EMD) was introduced, which intelligently separates jitter motion from the global motion vector (GMV) and demonstrates excel-lent performance in the majority of cases[9]. However, EMD may lead to in-correct mode separation due to mode mixing, necessitating more accurate algorithms for complex situations. To address this issue, several enhanced EMD methods

have been proposed such as ensemble EMD (EEMD) by Wu and Huang[10], but it has a disadvantage of incomplete removal of white noise added into EMD. Consequently, more advanced techniques called comple-mentary EEMD (CEEMD) and complete EEMD with adaptive noise (CEEMDAN) method were proposed[11][12]. Nonetheless, both CEEMD and CEEMDAN demand significant computational resources.
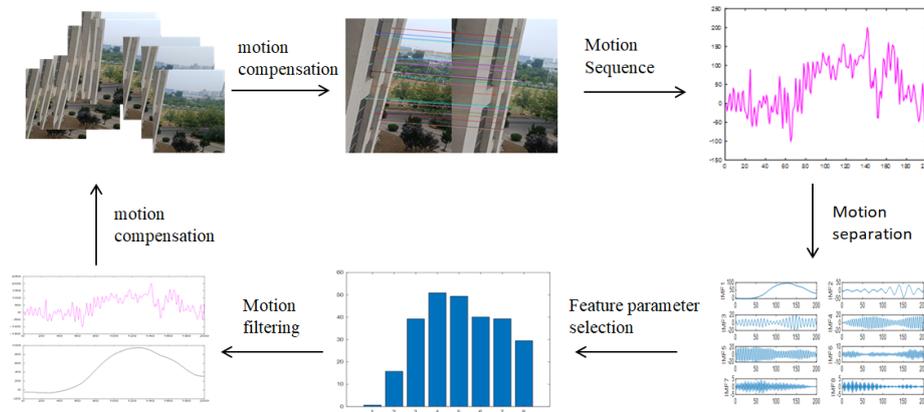

Figure 1: Digital Image Stabilization Framework Based on VMD.

This study introduces a DIS algorithm to tackle the video stabilization issue. Firstly, the global motion vector (GMV) is calculated using the SURF approach from a set of scouting video sequences[13][14]. Then the obtained GMV is decomposed using variational mode decomposition (VMD) to determine IMFs. Next, the scale factor analysis (SFA) is computed. Based on the SFA values, the boundary between modes is established for separating intentional motion from jitter motion successfully, enabling the reconstruction of both motions. The experimental outcomes demonstrate that the proposed method delivers smoother estimations with reduced errors compared to other algorithms.

The rest of this paper is organized as follows. Section 2 describes VMD and SFA. Section 3 illustrates the proposed DIS algorithm. Section 4 provides the experimental results and the comparison of the performances. Finally, a conclusion is drawn in Sec. 5.

2. Correlation principle

2.1. VMD Theory

The VMD approach is distinct from conventional recursive models, as it simultaneously explores the modes and their central frequencies. Through VMD, the signal can be separated into numerous band-limited modes u_k ($k = 1, 2,…, k$), where k is the number of modes. Each mode converges around the center frequency ω_k ($k = 1, 2,…, k$). Therefore, variational problem can be constructed, as shown by Equation (1):

$$\min_{\{u_k\},\{w_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]*e^{-jw_kt}\right\|_2^2\right\} \qquad s.t.\sum_k u_k = f \qquad (1)$$

where f is the input signal, δ is Dirac distribution, t is time script, and $*$ denotes convolution. To solve Equation (1), a quadratic penalty term α and Lagrangian multiplier λ are used to transform the constrained variational problem into the following unconstrained variational problem:

$$L\left(\{u_k\},\{w_k\},\lambda\right)=\alpha\sum_K\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]*e^{-jw_kt}\right\|_2^2+\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle \qquad (2)$$

Then, using the alternate direction method of multipliers (by updating the u_k^{n+1}, ω_k^{n+1}, and λ^{n+1} alternately),the solution of the optimal problem can be obtained by searching the saddle point of Equation (2) . VMD is implemented as follows:

1) Initialize the modes u_k, center pulsation ω_k, Lagrangian multiplier λ and the maximum iterations N (5000 in this paper). The cycle index is set to $n = 0$;

2) The cycle is started, $n = n + 1$;

3) The first inner loop is executed, and u_k is updated according to following function:

$$u_k^{n+1} = \arg\min_{u_k} L\left(\left\{u_{i\prec k}^{n+1}\right\}, \left\{u_{i\geq k}^{n}\right\}, \left\{\omega_i^{n}\right\}, \lambda^n\right). \tag{3}$$

4) The second inner loop is executed, and ω_k is updated according to the following function:

$$\omega_k^{n+1} = \arg\min_{\omega_k} L\left(\left\{u_i^{n+1}\right\}, \left\{u_{i\prec k}^{n+1}\right\}, \left\{\omega_{i\geq k}^{n}\right\}, \lambda^n\right). \tag{4}$$

5) λ is updated according to the following:

$$\lambda^{n+1} = \lambda^n - \tau\left(f - \sum_k u_k^{n+1}\right) \tag{5}$$

6) Steps (2)–(5) are repeated until convergence, as follows:

$$\sum_k \frac{\|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} \prec \varepsilon \tag{6}$$

where τ is an update parameter, ε is a small number (0.01 in this paper). The solution to update u^k and ω_k can be solved in the spectral domain, as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i\neq k}\hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{7}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 \, d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 \, d\omega} \tag{8}$$

After obtaining the modes in the frequency domain, they are converted into the time domain via inverse Fourier transform. Based on Dragomeretskiy's theory, two critical parameters that affect the outcome are the penalty α parameter and the number of modes k.

First, Dragomeretskiy suggested that if the principle frequencies of the sub-components are estimated a priori, then a low α is preferred to use because ω_k gains freedom of mobility to the appropriate modes. In the proposed method, a low α (100) is preferred because no prior frequencies of the sub-components are given. Secondly, when α is small, there are two possible issues. Either one mode is shared among neighboring modes (underbidding), or several additional modes comprise mostly texture with a weak structure (overbidding). In the former case, intentional motion and jitter motion may exist in the same mode, leading to poor results. If excessive modes are decomposed, it may not significantly improve performance but require more computational resources. In our simulation, we found that setting the number of modes to 8 meets most testing requirements.

2.2. Sampling fluctuation analysis

Sampling fluctuation analysis is a method to measure signal volatility by calculating sampling loss area. Consider signal $x(t)$, the signal $x_1(t)$ is obtained by interval sampling of $x(t)$. The sequence $x_2(t)$

has the same length with $x(t)$, which is obtained by linear interpolation of $x_I(t)$. as shown in Figure 2. The yellow area between the two curves is the sampling loss area. The smaller the sampling loss area, the more consistent the sampling waveform is with the original waveform, and the smaller the sequence volatility. On the contrary, the larger the sampling loss area, the more dramatic the signal fluctuation. The calculation formula is as follows:
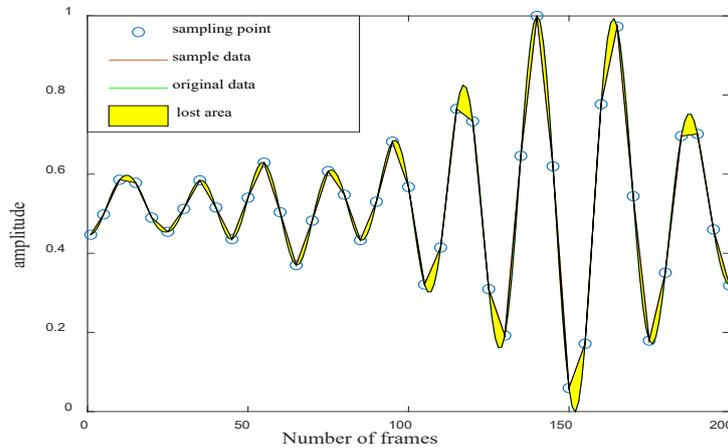
$$SL = \sum_{i=1,2,...,N} abs\left[A(t) - B(t)\right]$$

(9)

Figure 2: sampling loss area of a noisy signal

3. Proposed Digital Image Stabilization Method

The proposed DIS framework involves three essential steps: motion estimation, motion separation, and intentional motion vector reconstruction. In the first step, the GMV is estimated using the SURF matching algorithm. Subsequently, VMD is used to decompose the GMV into different modes, followed by SFA to identify the intentional and jitter motion modes for separating them. In the final stage, the sum of jitter motion modes yields the jitter motion vector, while the difference between the resulting sum and the original GMV gives the intentional motion vector. Figure 3 illustrates the framework of our proposed DIS method.
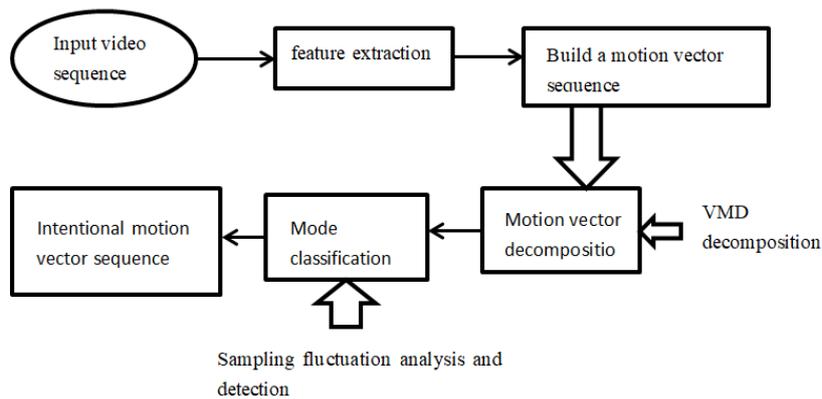

Figure 3: Digital image stabilization framework is proposed in this paper

3.1. Global Motion Vector Estimation

SURF feature point detection was proposed by Bay et al.[15]. By introducing integral image and box filter to extract feature points, the search process of feature points is optimized, the operation speed is fast, and it has good robustness under image scale and illumination changes. After extracting the SURF feature points of the two frames, the matching of the corresponding feature points is achieved by calculating the Euclidean distance between the two groups of feature points. The matching results of the SURF points of the given two images are shown in Figure 4, where the detected SURF feature points

and the corresponding matching results are shown.

Figure 4: Schematic diagram of SURF feature point matching

The motion parameter matrix obtained by the above motion estimation includes three forms of motion: translation, rotation and scaling.Therefore, the global motion parameter matrix is a time-varying matrix, and the size of the matrix elements is the amplitude of the three motions in the field of view.Take the video shown in Figure 5 as an example, 200 frames of images are captured for calculation. The horizontal translation, vertical translation and rotational motion of the video are shown in Figure 6.

Figure 5: Testing Video
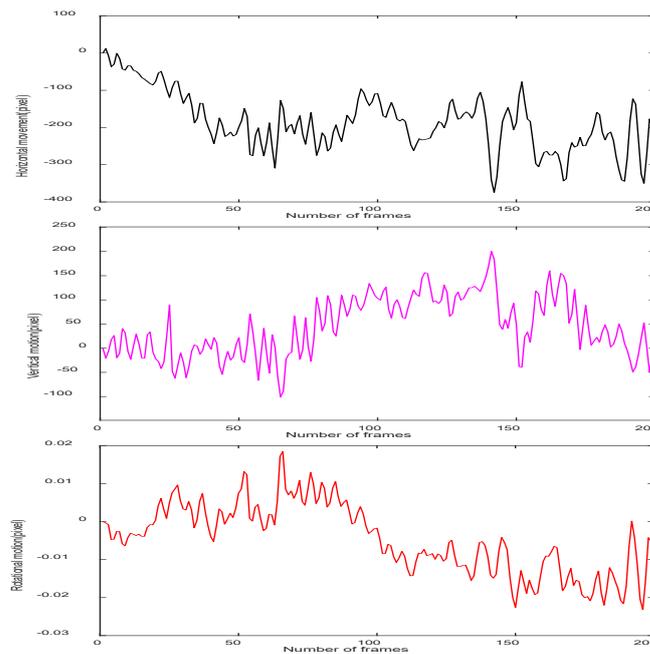
Figure 6: Horizontal, vertical, and rotational motion of 200 frames

3.2. Digital Image Stabilization Motion Separation Method

The procedures for different type motion are the same. Thus, takes horizontal motion as example for demonstration in our method, although both vertical, horizontal and rotational motions are present. After motion estimation, the obtained GMV of the video sequence is considered a time-varying signal $G(t)$. The amplitude of $G(t)$ can be regarded as the position of the camera. $G(t)$ contains both intentional motion and jitter motion. VMD is applied to $G(t)$ to separate the two types of motion.

Consider the following signal.

$$G(t) = x(t) + n(t) \tag{10}$$

where $G(t)$ is the original signal, $x(t)$ is the targeted signal, and $n(t)$ is the noisy signal, as shown in Figure 7. After the original signal is decomposed via VMD, we obtain eight modes (the last mode is the residue), as shown in Figure 8. This figure illustrates that the modes are arranged from low to high frequency.
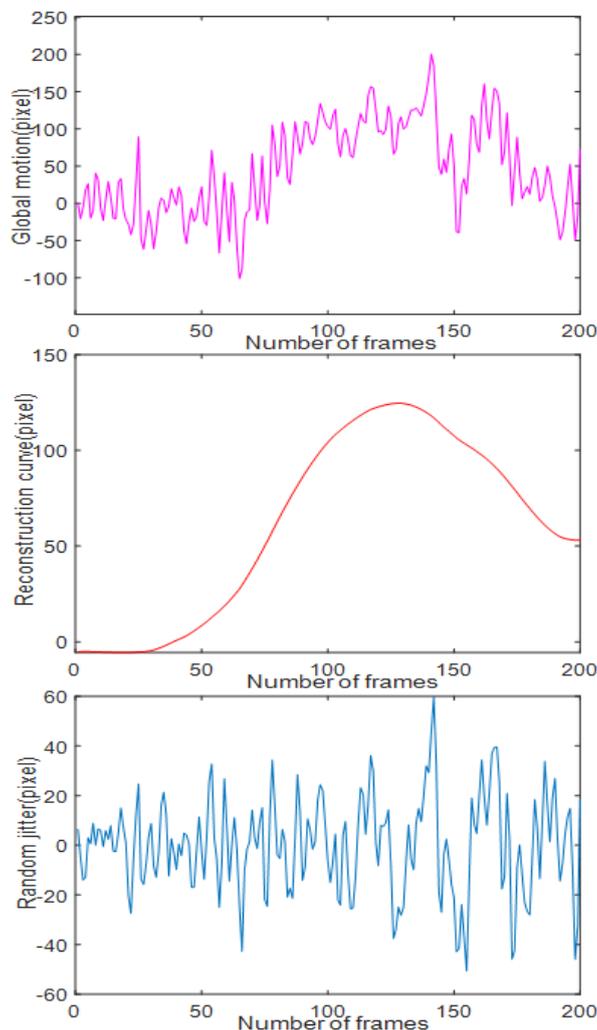

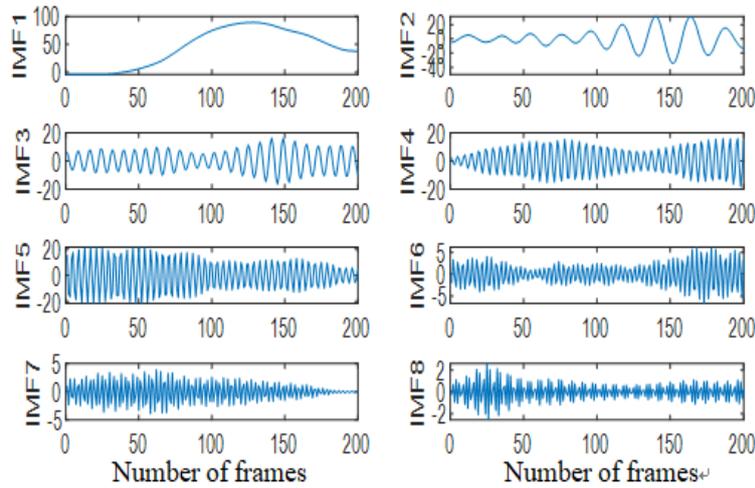*Figure 7: Results of contrast between original GMV and reconstructed intentional motion, random jitter*

Figure 8: VMD decomposition results of vertical motion

The proposed filtering method is based on the concept of enhancing the coherency of the modes. Following the theory of VMD, the relationship between Intrinsic Mode Functions (IMFs) and the original signal is as follows:

$$G(t) = \sum_{i=1}^{m} IMF_i + \sum_{i=m+1}^{n} IMF_i \tag{11}$$

Where m is the index of the relevant mode that separates targeted modes from the signal and n is the number of modes. The noise signal modes are found from the m-th mode to the n-th mode, whereas the remaining modes are targeted signal modes.

$$n(t) = \sum_{i=1}^{m} IMF_i \tag{12}$$

$$x(t) = \sum_{i=m+1}^{n} IMF_i = G_{(t)} - n_{(t)} \tag{13}$$

In general, the first mode of VMD contains low-frequency and high-amplitude components, which can be considered a targeted signal mode. To identify the relevant mode, one can evaluate if there is a significant change in the mode's structure. In our study, we calculated the Smoothed L1-norm Amplitude (SLA) of each mode. If the SLA is low, it indicates that the mode belongs to the targeted signal. Therefore, the relevant mode can be identified based on the SLA values. By contrast, SLA will demonstrate a sharp increase and stay at high level when two modes belong to noisy signal. That is, when SLA suddenly increases a biggest value, index m of IMF will be used to separate the noise and targeted signal modes. When the SLA of a mode is low, it suggests that the mode before Intrinsic Mode Function (IMF) decomposition is dominated by the targeted signal, while the other modes are comprised of noisy signal components. Figure 9 illustrates this concept, where the first bar represents the SLA of the first mode, the second bar represents the SLA of the second mode, and so on. By examining the SLA values, we can identify the modes that are predominantly composed of the targeted signal and those that contain primarily noise. The first SLA is the lowest, and the increase of the second SLA is the biggest. Thus, the first mode is the only targeted signal mode.

According to the above analysis, the proposed inter-frame motion vector reconstruction process is as follows:

1) GMV sequence is decomposed into a series of IMFs by VMD method;

2) Calculate the iterative loss area of each IMF;

3) Select the critical IMF according to the sampling volatility analysis value of each IMF to determine the position of the cut-off point between random jitter and intentional movement;

4) Intentional motion vector sequence and random jitter vector sequence can perform additive reconstruction of the corresponding IMF.
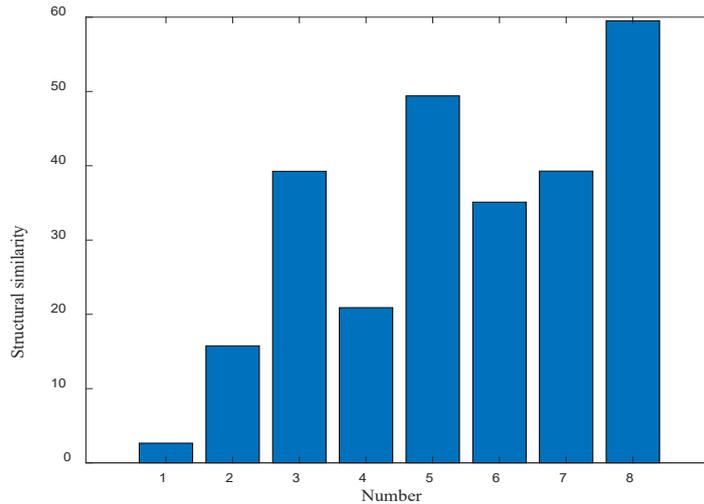
Figure 9: Loss area of each mode

4. Experimental comparison of digital stabilization results

4.1. Performance of VMD

To illustrate the effectiveness of VMD, several tests are performed to evaluate mode separation performance. Given a known clean signal $fh(t)$, contaminated the signal with coloured noise as shown in Figure 10.

$$f(t) = fh(t) + n(t) \tag{14}$$

Where $n(t)$ is the coloured noise signal, which consists of many signals with different frequencies and high amplitude. The classic EMD algorithm, CEEMDAN and the VMD algorithm were used to decompose. The decomposed results are shown in Figure 11, Figure 12 and Figure 13 respectively, and the runtime are shown in Figure 14. It can be seen from the figure that the first IMF of the classical EMD decomposition contain distinct different frequency components (mode mixing). If this phenomenon occurs, the decomposed IMF may contain both intentional motion and random jitter, and it will not able to separate the intentional component from the jitter component. On the contrary, CEEMDAN and VMD can effectively avoid mode mixing problem, and each IMF is a narrowband signal, which do a better job of separating deeper and removing mutually affecting. Additionally, in contrast with the CEEMDAN, the VMD has the distinct advantages of being short in operation time, which indicates that the VMD outperforms the EMD and its improved version.
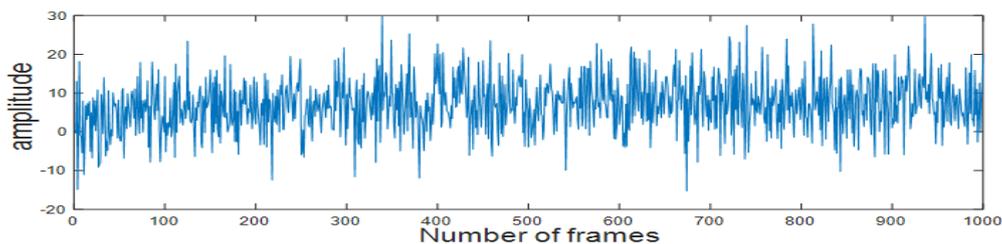

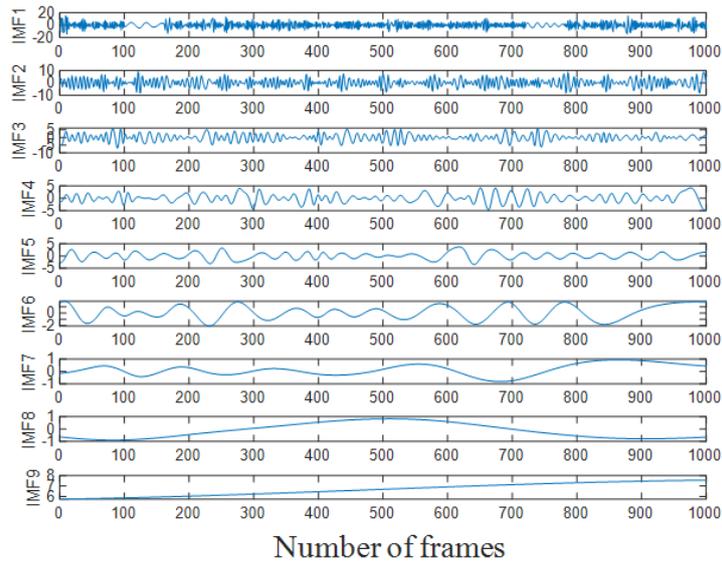*Figure 10: This caption has one line so it is centered.*
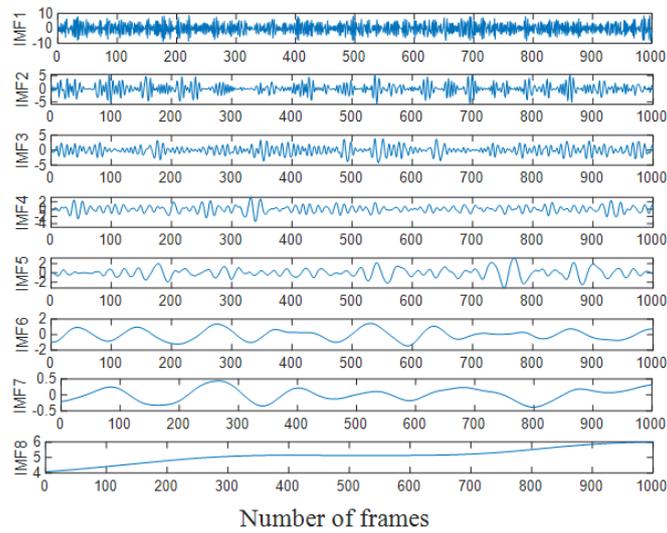
Figure 11: EMD decomposition results

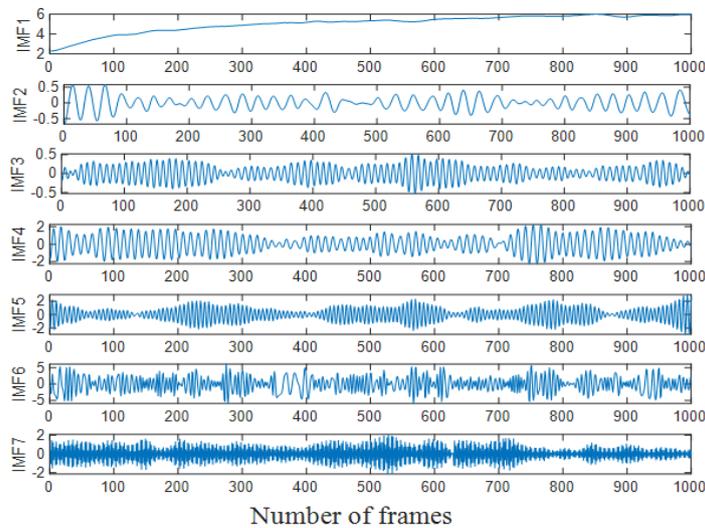*Figure 12: CEMDAN decomposition results*

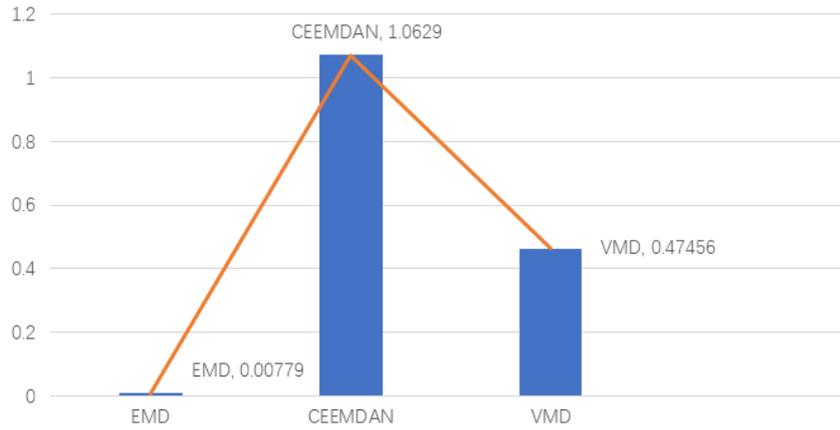*Figure 13: Results of VMD decomposition*

Figure 14: Time comparison char

4.2. Performance of the VMD-SFA method in DIS

To evaluate the effectiveness of the proposed VMD-SFA method, several simulation tests were conducted. The experiments were performed using a camera mounted on a holder mechanism, as demonstrated in Figures 15-17. In this paper, we employed two metrics, namely Root Mean Square Error (RMSE) and Peak Signal-to-Noise Ratio (PSNR), to quantify the performance of our proposed method [16]:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2} \tag{15}$$

$$PSNR(i) = 10\lg\frac{255^2}{MSE(i)} \tag{16}$$

We conducted experiments on three typical unstable scouting video sequences. In the first test, the jitter motion followed a Gaussian distribution with a fixed mean and variance. In the second test, the jitter motion obeyed a uniform distribution. For test 3, the jitter motion obeys a Gaussian distribution and the amplitude is time-varying. The root mean square errors (RMSE) between the filtered signal and the original signal[17] are calculated, where y_i is the intentional motion curve decomposed and reconstructed after filtering, and $\hat{y}_i$ is the original signal. m is the number of frames.. The lower the RMSE, the better the filtering effect. We conducted the experimental tests using MATLAB R2018b on a PC equipped with a 1.19GHz Intel Core i5-1035G1 CPU and 16GB RAM. Figures 15-17 depict three groups of images that were extracted from different video sequences. The actual Global Motion Vector (GMV), ground truth intentional motions, and retrieved intentional motions are shown in Figures 18-20. Table 1 summarizes the RMSE values obtained using four different DIS algorithms, including the MF[18], WD[19], enhanced EMD-based method[20], and the proposed method. The PSNR results between each adjacent two frames are shown in Figures 21-23.

Figure 15: Test Video 1

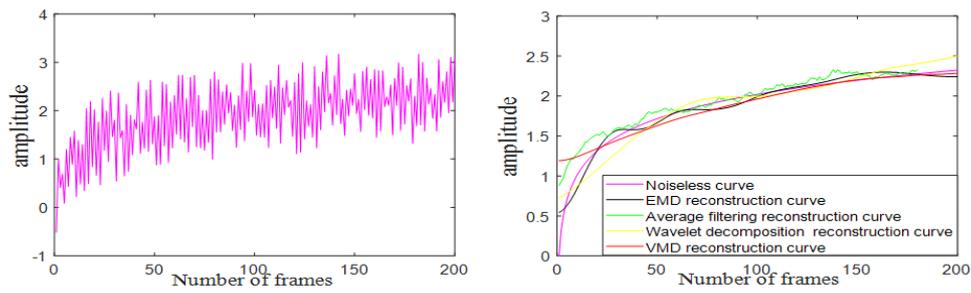*Figure 16: Test Video 2*

Figure 17: Test Video 3

Figure 18: Decomposition and reconstruction curve of simulation signals and each filtering method
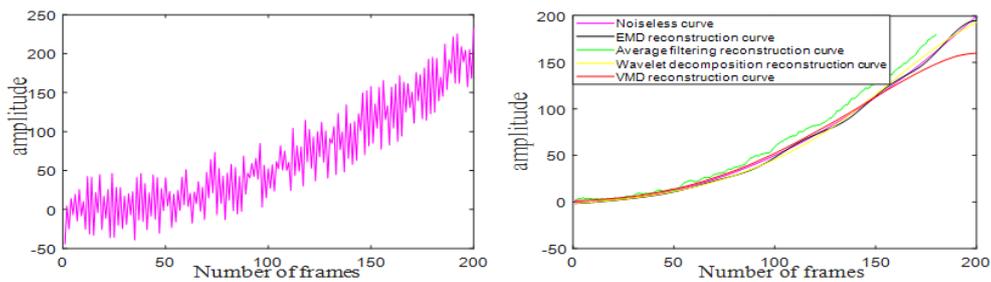

Figure 19: Decomposition and reconstruction curve of simulation signals and each filtering method
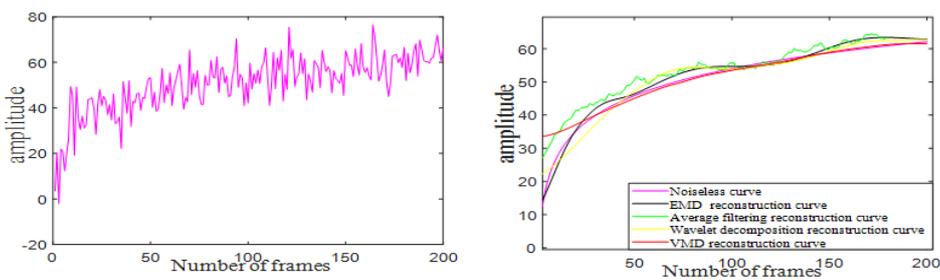

Figure 20: Decomposition and reconstruction curve of simulation signals and each filtering method

Table 1: Reconstructed signal and non-noise signal RMSE values by different filtering methods.

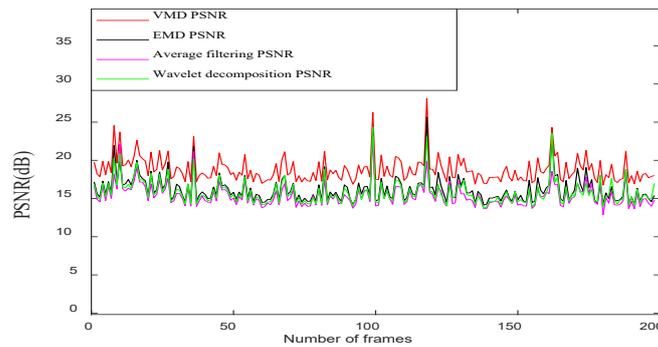| Method	RMSE		
	test.1	test.2	test.3
Average filtering	0.1305	6.0621	3.1322
Wavelet Decomposition	0.1106	4.2956	3.1244
EMD Decompose	0.1112	3.2350	2.8298
VMD Decompose	0.1078	3.0678	2.6497

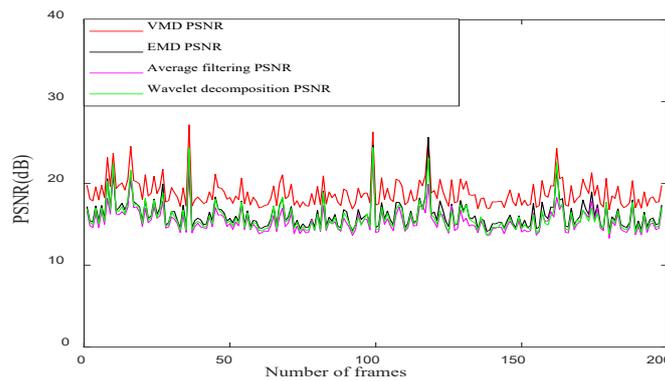*Figure 21: Testing the PSNR of Video 1.*

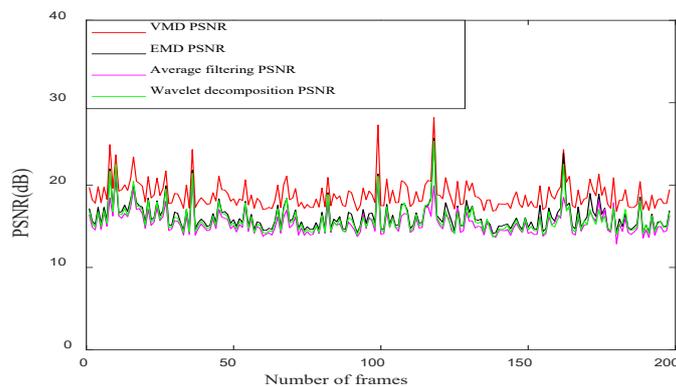*Figure 22: Testing the PSNR of Video 2.*

Figure 23: Testing the PSNR of Video 3.

Based on the results presented in Table 1, we can conclude that the Motion Flow (MF) algorithm generated the poorest outcomes, exhibiting unstable performance. The performance of MF is highly dependent on the window size parameter [21]. A larger window size produces a smoother intentional motion vector, while a smaller window size may lead to higher noise levels. In our study, we set the window size to 21, which is suitable under certain conditions but may not be optimal for other scenarios. On the other hand, the Wavelet Decomposition (WD) method may struggle to select an appropriate

wavelet basis function that is applicable in all conditions [22]. The performance of the WD and MF methods may improve with a well-selected basis function but may still be limited in their applicability to changing conditions, particularly in complex vehicle-mounted DIS systems. Comparing traditional methods with mode decomposition methods, we can generally conclude that mode decomposition methods tend to perform better. However, it is worth noting that the Empirical Mode Decomposition (EMD) method outperformed other modes in Tests 2 and 3. This result can be attributed to the difficulty of disentangling intentional and jitter motions due to the overlap of frequency information (mode mixing) in complex conditions. The proposed VMD-SFA method consistently produced the lowest RMSE values and the highest SNR values across all tests, demonstrating its effectiveness in addressing jitter noise stabilization in scouting video sequences. And the PSNR curve stay at the highest level in all of these methods.

5. Conclusions

This study proposed a DIS method based on VMD and SFA. GMV is estimated using SURF feature point matching algorithm. Then, GMV is decomposed via VMD. According to the SFA values, relevant modes of intentional and jitter motions are determined. Performance of the proposed method is compared with several state-of-the-art methods. Simulation results show better performance of the proposed method than other related methods based on quantitative comparisons of RMSE and PSNR values.

References

[1] Zhang Ning, Yang Yuan, Wu Jianhua, Zhao Ziqian, Yin Guodong. Vehicular Electronic Image Stabilization System Based on a Gasoline Model Car Platform [J]. Chinese Journal of Mechanical Engineering, 2022, 35(1).

[2] Qin Shumin, Zhao Lizhen, Tian Xingke, Li Lidan, Liu Shuyun. Research on Electronic Image Stabilization Technology of Vehi-cle-Mounted Remote-Controlled Weapon Station[J]. Journal of Physics: Conference Series, 2021, 2083(3).

[3] N. Zhang, H. B. Zhang, J. H. Wu, Y. Yang, G. D. Yin. An Electronic Image Stabilization Algorithm for Vision System of Intelligent Vehicles [J]. Journal of Tongji University, 2022, 50(04):497-503.

[4] Mingyuan Fan, Ziyuan Tong, Shoufeng Tang, Minming Tong, Bo Wang. Research of Digital Image Stabilization on Airborne Video [J]. IOP Conference Series: Materials Science and Engineering, 2018, 381(1).

[5] Duo Hao, Qiuming Li, Chengwei Li. Digital Image Stabilization Method Based on Variational Mode Decomposition and Relative Entropy [J]. Entropy, 2017, 19(11).

[6] Meguro M, Taguchi A, Hamada N. Data-dependent weighted average filtering for image sequence restoration [J]. Electronics and Communicat-ions in Japan (Part III Fundamental Electronic Science), 2001, 84 (4): 1-10.

[7] J. L. Wang. Wavelet analysis of volatile organic compounds in the atmosphere [J]. Cleaning World, 2022, 38(09):60-62.

[8] B. You, C. F. Zhang. Image denoising based on wavelet adaptive threshold and bilateral filtering [J]. Computer Engineering and Design, 2019, 40(08):2278-2282.

[9] K. Ioannidis and I. Andreadis, "A digital image stabilization methodbased on the Hilbert– Huang transform," IEEE Trans. Instrum. Meas. 61(9), 2446– 2457 (2012).

[10] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method, " Adv. Adapt. Data Anal. 1, 1– 41(2009).

[11] J. -R. Yeh, J. -S. Shieh, and N. E. Huang, "Complementary ensembleempirical mode decomposition: a novel noise enhanced data analysismethod," Adv. Adapt. Data Anal. 2, 135– 156 (2010).

[12] M. E. Torres et al., "A complete ensemble empirical mode decompo-sition with adaptive noise," in Proc. of 2011 IEEE Int. Conf. on Acoustics, Speech and Signal, pp. 4144– 4147 (2011).

[13] Y. Y. Zou, Z. L. Yang, L. D. Yang. Image matching algorithm based on improved SURF [J/OL] Computer Systems & Applica-tions: 1-7[2022-12-04].

[14] Li Xiaoguang, Zhu Juan, Ruan Yiming. Vehicle Seat Detection Based on Improved RANSAC-SURF Algorithm [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2021, 35(05).

[15] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (surf) [J]. Computer Vision & Image Understanding, 2008, 110 (3): 346-359.

[16] Ahmed Aldahdooh, Enrico Masala, Glenn Van Wallendael. Framework for reproducible objective video quality research with case study on PSNR implementations [J]. Digital Signal Processing. June 2018, 77: 195-206.

[17] Kamble V. B.; Deshmukh S. N. Comparision between Accuracy and MSE, RMSE by Using Proposed Method with Imputation Technique. Oriental Journal of Computer Science and Technology. 2017, 10: 773-779.

[18] Cui, J . W.;Wang, D. Q. ;Liu. J. Y.; Single Image Dehazing Algorithm Based on Average Filtering and Wavelet Transform. Com-puter & Digital Engineering. 2022(06).

[19] Shaikh, W A. Shah S F, et al. Pandhiani Siraj Muhammad; A hybrid forecasting model based on the group method of data handling and wavelet decomposition for monthly rivers streamflow data sets. Open Physics. 2022. PP 1096-1111

[20] Markus G; Patrik Z, et al. Improved EMD-based Oscillation Detection for Mechatronic Closed-Loop Systems. IFAC-Papers On Line 2019. PP 370-375.

[21] Wang, K J. Xiong, X Y. Ren, Z. Highly efficient mean filtering algorithm. Application Research of Computers. 2010(02).

[22] Q. Wang, X. Chen. Selection of wavelet bases in wavelet analysis of microseismic signals of rock fractur [J]. Electronic Design Engineering., 2016, 24(21):126-128+131.