PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

An objective assessment method for video stabilization performance

Wang, Yiming, Huang, Qian, Sun, Sidong, Ye, Feng, Wang, Yanfang

Yiming Wang, Qian Huang, Sidong Sun, Feng Ye, Yanfang Wang, "An objective assessment method for video stabilization performance," Proc. SPIE 11179, Eleventh International Conference on Digital Image Processing (ICDIP 2019), 111792M (14 August 2019); doi: 10.1117/12.2540340

SPIE.

Event: Eleventh International Conference on Digital Image Processing (ICDIP 2019), 2019, Guangzhou, China

An Objective Assessment Method for Video Stabilization Performance

Yiming Wang[a], Qian Huang[a, b]*, Sidong Sun[a], Feng Ye[a, c], Yanfang Wang[a]
[a] College of Computer and Information, Hohai University, No. 8 Focheng West Road, Jiangning District, Nanjing, China, 211100; [b] Nanjing Huiying Electronics Technology Corporation, No. 33 Dongqi Road, Jiangning District, Nanjing, China, 211103; [c] Nanjing Longyuan Micro-Electronics Company, No. 68 Shengtai Road, Jiangning District, Nanjing, China, 211100

ABSTRACT

The intent of video stabilization to remove violent jitter in the video and reduce video distortion. Recently, many excellent video stabilization algorithms have been proposed. However, it is a question to assess the video stabilization performance objectively. In order to solve the problem, we propose a novel objective assessment method for video stabilization performance by computing the distance, distortion and similarity between the stabilized and reference frames. According to published video stabilization database, data sets are built and tested. Finally, the experimental results demonstrate the accuracy of our proposed assessment method and the results are consistent with subjective human eyes.

Keywords: assessment for video stabilization performance, distance, distortion, similarity

1. INTRODUCTION

With the development of electronic communication and multimedia technology, taking videos are becoming more and more popular in people's daily lives. However, in the process of shooting, Because of the professionalism of the photographer and the constraints of objective conditions, the videos taken are shaking violently, which makes the viewer feel tired and decreases the video quality. Video stabilization techniques have been studied from 1970s to remove unintentional motion, known as jitter and improve video quality. These stabilization schemes are divided into three categories: electronic image stabilization (EIS), optical image stabilization (OIS [1]) and digital image stabilization (DIS [2]). In recent years, digital video stabilization has been the focus of research. Digital video stabilization typically mostly adopts three-step methods: motion estimation, motion compensation and motion inpainting. According to the different motion models, it can be divided into 2D methods, 2.5D methods and 3D methods. Two-dimensional method [3-5] estimates the global motion of 2D affine motion parameters based on pixels or feature points, and then eliminates jitter and generates stable video by motion path smoothing or motion path fitting. In three-dimensional method [6], structure from motion (SfM) is used to restore the camera attitude and scene structure information, and motion compensation is carried out according to the restored information, to achieve stable video. Subspace constraints are implemented on feature trajectories and smoothed to stabilize it in 2.5-dimensional methods [7].

Nowadays, Video stabilization is an essential function in many hand-held devices. However, there is little objective assessment method for video stabilization, it is hard to evaluate the performance of video stabilization algorithm objectively. This paper proposes a novel mean to evaluate the performance of video stabilization algorithm objectively to make it easy for users to choose suitable video stabilization algorithms. The reminder of this paper is structured as follows: In section 2, the related work is viewed and full-reference and no-reference methods is compared. Section 3 introduces the novel objective assessment method and build video stabilization performance data sets. Section 4 demonstrates our experimental results and validate our proposed expression. Finally, Section 5 contains the conclusion of paper and our future's assignment.

2. RELATED WORK

Video stabilization quality evaluation [8-9] is divided into two categories: subjective and objective assessment methods. The subjective assessment method only involves the qualitative assessment made by the person. Human make a subjective qualitative assessment of the stabilization and jitter of the video. The method is established in a statistical sense. In order to ensure that the subjective assessment of the image is statistically significant, the numbers of observers should be sufficiently. A mean opinion score (MOS) [10] is a commonly used measure for video stabilization quality

Eleventh International Conference on Digital Image Processing (ICDIP 2019) edited by
Jenq-Neng Hwang, Xudong Jiang, Proc. of SPIE Vol. 11179, 111792M © 2019
SPIE · CCC code: 0277-786X/19/$21 · doi: 10.1117/12.2540340

assessment. But, Achieving MOS ratings may be slow and expensive. So, it is not suitable for the stabilization assessment of large number data sets.

The objective assessment method establishes a mathematic model based on the subjective visual system of the human, and calculates the stabilization through a specific expression. It is subdivided into full reference [11] and no reference [12] methods. At present, Full reference method is mainly performed in objective assessment method. Full reference method requires a stable ground truth video as a reference. A stable ground truth video is combined with jitter to create jitter video. Then, stabilize it and compare the obtained stable video and the ground truth video to get the result of the assessment. Most common full reference method is peak signal to noise ratio (PSNR), mean square error (MSE) [13] and Structure Similarity (SSIM) [14]. They are widely adopted because of their applicability and computability. Nowadays, no reference methods develop quickly in video stabilization. Wu el al [15] proposed Reconstruction-based No-Reference Video Quality Assessment. Liu el al [12] introduced cropping size and geometrical distortion to demonstrate the merits of their method. But, the usability of the method is poor.

Compared with the non-reference quality assessment method, the full reference quality assessment method is simpler and more applicable. In this paper, a novel method is proposed to increase the result of Hui Qu's paper [14]. Considering the video distortion and human visual factors in the process of video stabilization, we refine the distance between the smooth trajectory and employ partly PSNR and SSIM to improve the consistency between the results of objective assessment algorithm and subjective effect. A large number of experiments also verify the effectiveness of the proposed full-reference video stabilization performance assessment algorithm.

3. OBJECTIVE ASSESSMENT METHODS

In this paper, jitter is randomly added to the ground truth video to generate jitter video. The ground truth stable video is compared with the stabilized video which was originally shaking, and then the stabilization is assessed. Due to the lack of experimental equipment and time, we approximately see the video which are in [16] as a pair of jitter video and the ground truth video, contain 144 videos. The data set was divided into seven types: simple, rotation, zooming, parallax, driving, crowed and running. Four types of videos are selected. The purpose of this paper is to confirm the performance of our algorithm. Later, the ground truth videos are collected to further validate our algorithm.

This paper assesses the performance of several stabilization algorithms from three aspects: camera motion trajectory, video similarity and video distortion. PSNR and SSIM [17] are introduced to judge the quality of stabilized video and the ground truth. Among them, PSNR [18-19] cannot reflect subjective feelings, SSIM calculation is slightly complicated, but its value can better reflect subjective feelings. Objective assessment formula is defined as:

$$\begin{aligned} D_{mv} &= \sum_{i=1}^{N} \frac{1}{\sqrt{(x(i)_{gmv} - x(i)_{smv})^2 + (y(i)_{gmv} - y(i)_{smv})^2}} \\ &+ \lambda 10 log_{10} \frac{255^2}{\frac{1}{N}\sum_{i=1}^{N}|GMV(n) - SMV(n)|^2} \\ &+ \mu \frac{1}{N} \sum_{i=1}^{N} \left| ssim_{gmv}(n) - ssim_{smv}(n) \right| \end{aligned} \tag{1}$$

where Euclidean distance between two points represents the trajectory of per frame of the ground truth video and the stabilized video. N represents the number of frames. $GMV(n)$ denote n^{th}GMV of the reference video sequences and $SMV(n)$ denote n^{th}SMV of the smoothed video sequences. $ssim_{gmv}(n)$ and $ssim_{smv}(n)$ represents n^{th} ssim value of reference and smoothed video sequences respectively. And λ, μ are the weights in equation.

In order to verify the effectiveness of our proposed algorithm, we take four types of videos: simple motion, rotation motion, parallax motion and running motion.

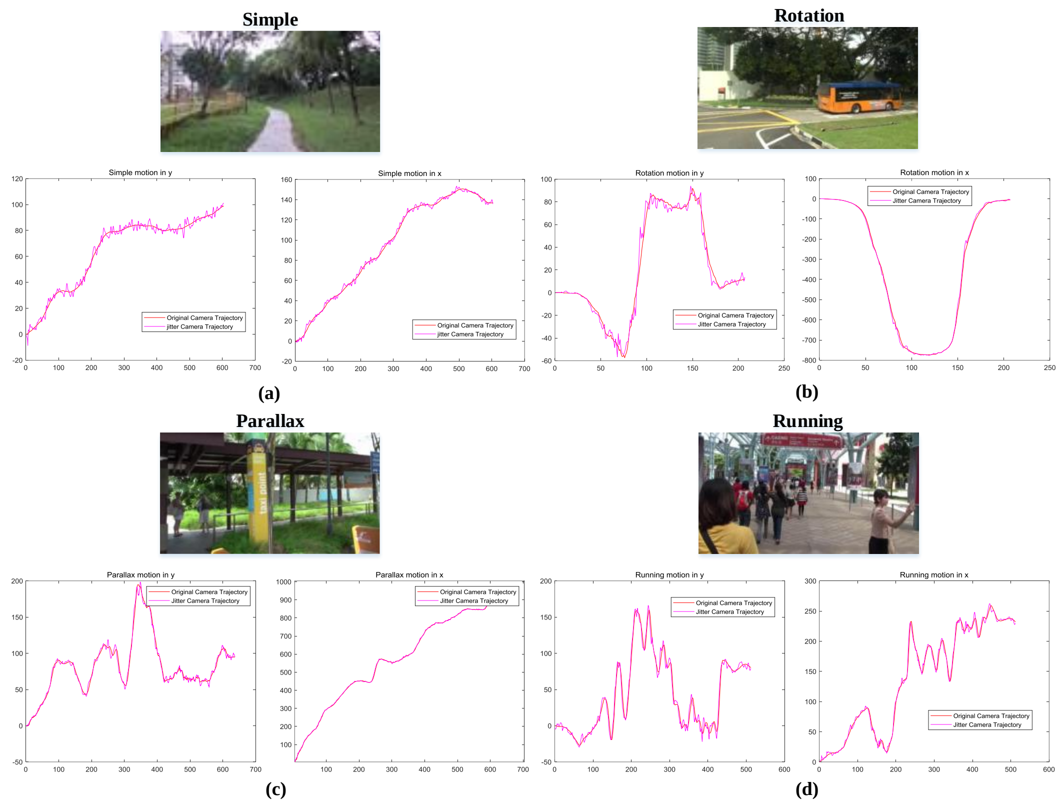

Figure 1. Jitter video and the original video (the ground truth) trajectory.

Fig. 1(a) shows the simple video on campus paths. Fig. 1(b) shows the large rotation on the bus. Fig. 1(c) shoots the video which contains the large parallax. And Fig. 1(d) shows that the video was taken while running. At the same time, Fig. 1 (a), (b), (c) and (d) shows the comparison of original camera and jitter camera trajectory. Details of data set (e.g. jitter severity, similarity) are presented in Table 1.

Table 1. Motion in test sequences.

motion	frame numbers	direction	severity	similarity
simple	616	x, y	severe	yes
rotation	217	x, y	light	no
parallax	639	x, y	middle	no
running	519	x, y	middle	yes

4. EXPERIMENTAL RESULTS

We have a data set which contains the ground truth, jitter video and stable video obtained by processing jitter video using various video stabilization algorithms. The experiment consists of three parts: (1) compares the objective assessment method with the subjective results to verify the effectiveness of our proposed algorithm. (2) the outcome of our proposed assessment is compared with Hui Qu's paper to further prove the superiority of our algorithm. test environment for all experiments is windows 10 platforms, 64-bit computer operating system with an Intel Core i7 2.60 GHz processor and 16GB of RAM.

In order to ensure the effectiveness of our algorithm, three kinds of video stabilization software are selected on the market: Virtual Dub Deshaker [20], Adobe After Effects (AE) Warp Stabilizer VFX [21]**,** Google YouTube VideoProc[22]. In fact, AE Warp Stabilizer VFX is based on subspace algorithm [7], Deshaker adopts the stabilization algorithm which contains motion estimation based on block matches and motion compensation based on bicubic interpolation. YouTube VideoProc mainly realizes the L1 norm optimization of adjacent frames [23].

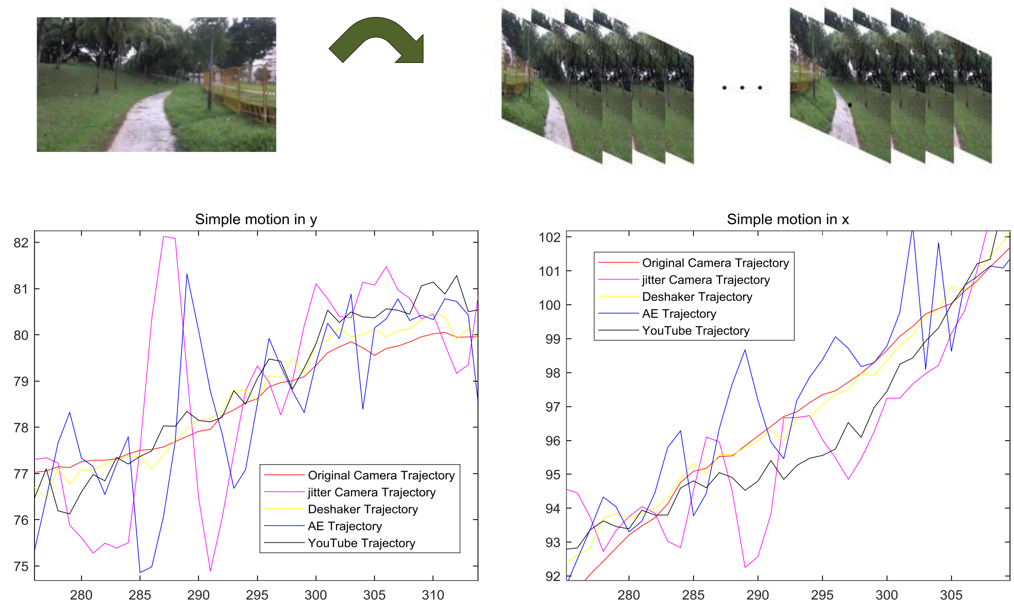

Figure 2. Paths of 'simple' in x and y direction: original camera trajectory is the camera motion of ground-truth video. Other paths are jitter, Deshaker, AE and YouTube trajectory.

Figure 3. Frames of ground-truth video "simple 19", its shaking video and videos for three stabilized algorithms. The first row is frames in jitter video with 640×360, the second row is corresponding frames in ground-truth video with 640×360, the third, fourth and bottom rows are Deshaker, AE and YouTube processed video frames with 640×360.

In Qu's [14] paper, he compared the distance of each frame between paths and the SSIM between consecutive frames respectively. But video stabilization performance assessment should be determined mainly by video stability, distortion and subjective perception of the human. So, the distance, PSNR and SSIM are combined. These results can also be reflected in Fig. 2.

Table 2. D_{mv} per frame between stabilized videos and reference videos for three algorithms.

video type	video name	stabilized and reference videos for three algorithms			
		Shaking	Deshaker	AE	YouTube
simple camera	simple	5.71	**6.84**	6.26	5.84
rotation camera	rotation	2.50	5.21	4.92	**5.71**
	parallax	2.39	3.84	**4.50**	3.71
running camera	running	5.38	**6.30**	5.98	6.10

Experiments show that the values of weights λ, μ are equal to three and one in equation separately and the result is more precise. From simple camera in the Table 2, we can observe the result of Deshaker algorithm is better than other algorithms in simple and running camera motion. The result of AE algorithm is the best stabilized algorithm in rotation parallax and result of YouTube algorithm is higher than others in camera motion with parallax.

Simple motion is chosen for the subjective observation. Motion curves in simple camera motion are shown in Fig. 2. Fig. 3 compares five successive frames 300, 301, 302, 303 and 304 in the testing sequences. By comparing five consecutive frames the motion paths and image quality are seen. Meanwhile, we also watch motion curve of different algorithms in fig2. Deshaker trajectory is the closest to the ground-truth trajectory in three algorithms. According to curves and data in the Fig. 2, Fig. 3 and Table 2, the objective assessment method is consistent with subjective results, our objective video stabilization assessment can be validated. Qu's paper compared the distance of each frame between paths and the SSIM between consecutive frames respectively. However, the distance, PSNR and SSIM are combined in our proposed expression. Finally, the experiment proves the superiority of our algorithm.

5. CONCLUSION

We propose a new objective assessment method for video stabilization performance. In this paper, new expression is adopted to combine the distance, PSNR and SSIM. The database in[1] is selected. Four types of videos are selected to validate the accuracy of our formula. Furthermore, simple motion is described by curves. Based on these videos, three kinds of video stabilization software on the market are assessed by our proposed algorithm. Experiments show that the effectiveness of our assessment method and is consistency with the subjective assessment results of human eyes.

With pairs of ground-truth and shanking videos, it is possible for us to make our objective assessment for video stabilization performance more robust. And attributes between stabilized frames and reference frames are not just distances, PSNR and SSIM. In the future's work, we will explore more attributes to further improve the accuracy of objective assessment for video stabilization performance.

ACKNOWLEDGMENTS

This work is partly supported by the National Key Research and Development Program of China under grant no. 2018YFC0407905, the key research and development program of Jiangsu province under grant no. BE2016904, the Jiangsu water conservancy science and technology project under grant no. 2018057, the National Natural Science Foundation of China under grant no. 61300122 and 61502145, the 2017 Jiangsu province postdoctoral research funding project under grant no. 1701020C, the 2017 Six Talent Peaks Endorsement Project of Jiangsu under grant no. XYDXX-078, the fundamental research funds of China for central universities under grant no. 2018B679X14 and 2017B42214, and the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant No. KYCX18_0639.

REFERENCES

[1] M. Hansen, P. Anandan, K. Dana, G. van der Wal and P. Burt, "Real-time scene stabilization and mosaic construction," *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, USA, 1994, pp. 54-62.

[2] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation," in *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 1320-1325, Nov. 2003.

[3] Y. Matsushita, E. Ofek, Weina Ge, Xiaoou Tang and Heung-Yeung Shum, "Full-frame video stabilization with motion inpainting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1150-1163, July 2006.

[4] F. Zhang, J. Wang, H. Zhao, R. R. Martin and S. Hu, "Simultaneous Camera Path Optimization and Distraction Removal for Improving Amateur Video," in *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5982-5994, Dec. 2015.

[5] Y Wang, Q Huang, and D Zhang, "An image stabilization algorithm on camera trajectory estimation and feature block matching," in *Proceedings of the 10th International Conference on Internet Multimedia Computing and Service (ICIMCS '18)*. ACM, New York, NY, USA, August. 2018.

[6] J. M. Wang, H. P. Chou, S. W. Chen and C. S. Fuh, "Video stabilization for a hand-held camera based on 3D motion model," *2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, 2009, pp. 3477-3480.

[7] F. Liu, M. Gleicher, J. Wang, H.-L. Jin, and A. Agarwala, “Subspace video stabilization,” *ACM Trans. Graph.*, vol. 30, no. 1, pp. 4:1–4:10, Feb. 2011.

[8] S. B. Balakirsky and R. Chellappa, "Performance characterization of image stabilization algorithms," *Proceedings of 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, 1996, pp. 413-416 vol.2.

[9] C. Morimoto and R. Chellappa, "Evaluation of image stabilization algorithms," *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, Seattle, WA, USA, 1998, pp. 2789-2792 vol.5.

[10] S. Winkler, "Issues in vision modeling for perceptual video quality assessment." *Signal Processing* 78.2(1999):231-252.

[11] M. Niskanen, O. Silven and M. Tico, "Video Stabilization Performance Assessment," *2006 IEEE International Conference on Multimedia and Expo*, Toronto, Ont., 2006, pp. 405-408.

[12] S.-C. Liu, L. Yuan, P. Tan, and J. Sun, “Bundled camera paths for video stabilization,” ACM Trans. Graph., vol. 32, no. 4, pp. 78:1–78:10, Jul. 2013.

[13] M. J. Tanakian, M. Rezaei and F. Mohanna, "Camera Motion Modeling for Video Stabilization Performance Assessment," 2011 7th Iranian Conference on Machine Vision and Image Processing, Tehran, 2011, pp. 1-4.

[14] H. Qu, L. Song and G. Xue, "Shaking video synthesis for video stabilization performance assessment," *2013 Visual Communications and Image Processing (VCIP)*, Kuching, 2013, pp. 1-6.

[15] S. Yao, W. Lin, Z. K. Lu, E. P. Ong, M. H. Locke and S. Q. Wu, "Image Quality Measure using Curvature Similarity," *2007 IEEE International Conference on Image Processing*, San Antonio, TX, 2007, pp. III - 437-III - 440.

[16] http://liushuaicheng.org/SIGGRAPH2013/database.html

[17] A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," *2010 20th International Conference on Pattern Recognition*, Istanbul, 2010, pp. 2366-2369.

[18] M. Vranjes, S. Rimac-Drlje and K. Grgic, "Locally averaged PSNR as a simple objective Video Quality Metric," *2008 50th International Symposium ELMAR*, Zadar, 2008, pp. 17-20.

[19] Z. Wu and H. Hu, "Reconstruction-based no-reference video quality assessment," *2016 IEEE Region 10 Conference (TENCON)*, Singapore, 2016, pp. 3075-3078.

[20] http://www.guthspot.se/video/deshaker.htm

[21] https://www.adobe.com

[22] https://www.videoproc.com/video-process/youtube-video-stabilizer.html

[23] M. Grundmann, V. Kwatra and I. Essa, "Auto-directed video stabilization with robust L1 optimal camera paths," *CVPR 2011*, Colorado Springs, CO, USA, 2011, pp. 225-232.