Improved global motion estimation via motion vector clustering for video stabilization

Article *in* Engineering Applications of Artificial Intelligence · September 2016

DOI: 10.1016/j.engappai.2016.05.004

12 authors, including:

Andrey Valerievich Kopylov
Tula State University
44 PUBLICATIONS 145 CITATIONS

wei-chih Yeh
Chaoyang University of Technology
105 PUBLICATIONS 1,818 CITATIONS

Oleg Seredin
Tula State University
48 PUBLICATIONS 163 CITATIONS

Sy-Yen Kuo
National Taiwan University
462 PUBLICATIONS 6,305 CITATIONS

Engineering Applications of Artificial Intelligence 54 (2016) 39–48

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Improved global motion estimation via motion vector clustering for video stabilization

Bo-Hao Chen [a,*], Andrey Kopylov [b], Shih-Chia Huang [c,*], Oleg Seredin [b], Roman Karpov [b], Sy-Yen Kuo [d], K. Robert Lai [a], Tan-Hsu Tan [e], Munkhjargal Gochoo [f], Damdinsuren Bayanduuren [f], Cihun-Siyong Gong [g,*], Patrick C.K. Hung [h]

[a] *Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan*
[b] *Institute of Applied Mathematics and Computer Science, Tula State University, Tula, Russia*
[c] *Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan*
[d] *Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan*
[e] *Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan*
[f] *School of Information and Communications Technology, Mongolian University of Science and Technology, Ulaanbaatar, Mongolia*
[g] *Department of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan*
[h] *Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Ontario, Canada*

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form
7 May 2016
Accepted 9 May 2016

Keywords:
Video stabilization
Motion vector clustering
Shortest spanning path

ABSTRACT

Video stabilization technique is often used in handheld multimedia devices, whereas the difficulties in the accurate extraction aspect of global motion vectors restrict its development. This paper proposes a novel video stabilization approach that is based on the shortest spanning path clustering algorithm for effective and reliable estimation of the global motion vectors. As demonstrated in our experimental results, the proposed approach achieves superior stabilized effectiveness compared with the other state-of-the-art approaches based on both qualitative and quantitative measurements.

1. Introduction

Several types of handheld multimedia devices have been dramatically developed for the last several decades, such as mobile phones, tablets, and so on (Leu et al., 2012; Kherallah et al., 2009). These devices allow observer to acquire videos from anywhere. However, the video acquisition through the handheld multimedia devices usually suffers from annoying perturbations (e.g. unexpected image motion Pandian et al., 2013) caused by the observer's hand shaking.

In response, video stabilization techniques have played an essential role in handheld multimedia devices. The task of video stabilization techniques is to compensate the unwanted image motion and eliminate these annoying perturbations from video streams (Niskanen et al., 2006; Qu et al., 2013). Numerous video stabilization techniques have been proposed by which to improve video quality in the devices. In general, they attain video stabilization through conjunctive use of camera motion estimation, motion filtering, and motion compensation (Puglisi and Battiato, 2011). In particular, the camera motion estimation is the first essential process in the development of video stabilization techniques by which to provide stable effects of both motion filtering and motion compensation. Specifically, the video stabilization techniques are built on a key observation that the affine transform of the frames is caused by camera motion. According to this observation, the stable frames can be acquired by inverting the global affine transformation. Hence, these techniques can be divided into two major categories according to their capability to estimate the camera motion, that are intensity-based approaches (Puglisi and Battiato, 2011; Kwon et al., 2006; Chang et al., 2004; Ko et al., 1999) and feature-based approaches (Battiato et al., 2007; Yang et al., 2009; Litvin et al., 2003; Shen et al., 2009; Kang et al., 2012; Pinto and Anurenjan, 2011; Feng et al., 2013).

Video stabilization techniques belonging to the intensity-based approaches category directly employ the image textures as motion vectors in each frame of a video to estimate the global affine transform and then reconstruct the stable frames. For instance, Puglisi and Battiato (2011) employed a block-matching algorithm to collect numbers of motion vectors while combining with a voting strategy for detection of global motion vector from different spatial locations of a frame. In addition, the video

* Corresponding authors.
E-mail addresses: bhchen@saturn.yzu.edu.tw (B.-H. Chen), schuang@ntut.edu.tw (S.-C. Huang), alexgong@mail.cgu.edu.tw (C.-S. Gong).

http://dx.doi.org/10.1016/j.engappai.2016.05.004
0952-1976/

Fig. 1. FREAK descriptors from frame 361 and frame 362 in video sequence “Statue”. The lines across the symbol ○ and symbol + represent the connections of matched features. The solid and dotted black circles represent the incorrect-matched and inconsistent feature points, respectively.

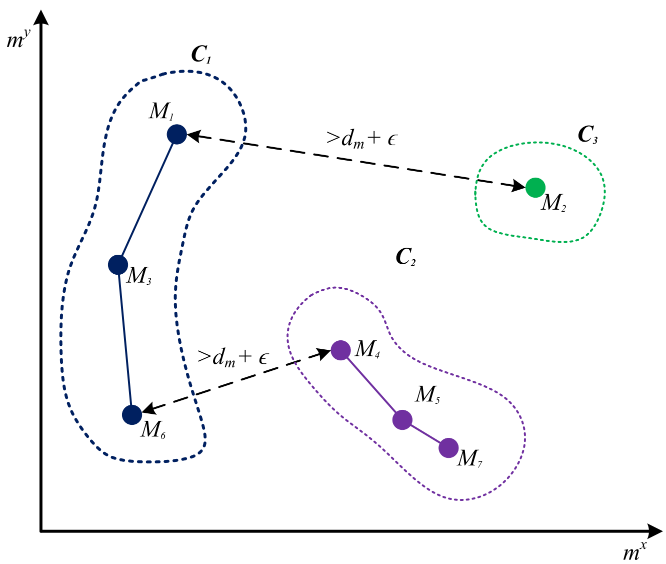

Fig. 2. Illustration of production process for each cluster.

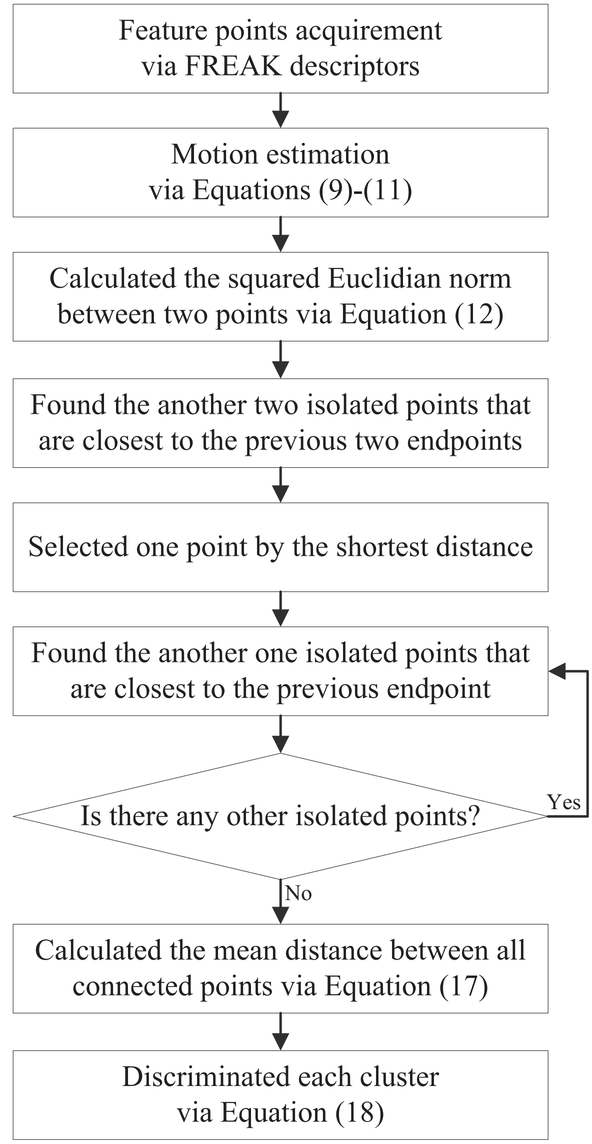

Fig. 3. Flowchart of proposed shortest spanning path based motion vector clustering.

stabilization technique of Kwon et al. (2006) was proposed, in which several motion vectors were estimated first by using the phase correlation-based motion estimation via four rectangular edge sub-images, after which the Kalman filter was used to extract the global motion vectors from those motion vectors. Chang et al. (2004) calculated the optical flows as global motion vectors based on brightness constancy assumption between adjacent frames and the camera motion was then estimated by fitting the optical flow field to a global affine motion model for stabilizing videos. Moreover, the video stabilization technique of Ko et al. (1999) obtained global motion vectors by using gray-coded bit-plane matching from those motion vectors realized by exploiting binary Boolean functions.

On the other hand, feature-based video stabilization approaches locate a sparse set of reliable features in adjacent frames for camera motion estimation. These features can be obtained from Scale Invariant Feature Transform (SIFT) (Lowe, 1999), Speeded Up Robust Features (SURF) (Bay et al., 2008), Kanade–Lucas–Tomasi feature tracker (KLT) (Shi and Tomasi, 1994), Fast Retina Key-point (FREAK) descriptors (Alahi et al., 2012), and so on. Hence, the global motion vectors can be estimated from these features by which to remove the unwanted image motion and further stabilize the videos. In the consideration of computational burden, the most widely adopted method for video stabilization is based on feature-based approaches (Matsushita et al., 2006). Hence, several methods based on the feature-based strategies have been popularly developed and implemented by many video stabilization applications.

The SIFT features are extracted by the method of Battiato et al. (2007); the motion vector integration technique is then utilized to filter the motion vectors produced from these SIFT features as global motion vectors for estimating camera movements. Moreover, Yang et al. (2009) also employed SIFT features to attain the motion vectors, whereupon the global camera motion was estimated by using the particle filters between successive frames. The method of Litvin et al. (2003) minimized the p-norm cost function to find the global motion vectors from the extracted feature points. In addition, Shen et al. (2009) proposed a feature-based video stabilization method that extracts features by using both the principal component analysis (PCA) and SIFT, after which the global motion vectors were detected by exploiting the RANdom SAmples Consensus (RANSAC) technique among the motion

Fig. 4. Comparison of the feature collections from frame 361 and frame 362 of the video sequence "Statue", with and without using the proposed method. (a) Collection result produced without using the proposed method, where the solid and dotted black circles represent the incorrect-matched and inconsistent feature points, respectively. (b) Collection result produced by using the proposed method, where both the incorrect-matched and inconsistent feature points are eliminated.

vectors of these PCA-SIFT features for camera motion estimation. The method of Kang et al. (2012) chose the global motion vectors by segmenting dominant planes that have similar depths after the features were produced by using KLT feature tracker. Pinto and Anurenjan (2011) employed SURF as stable feature points to be tracked, whereupon the heuristic modified trellis search technique was used to discriminate between local camera motion and global camera motion. A recent work proposed by Feng et al. (2013) built features by using the FREAK descriptors and further estimated the global camera motion from these features through the M-estimator Sample Consensus (MSAC) technique.

However, not all motion vectors obtained from each pair of features can offer accurate information in respect of the movement between the current and previous frames. There are two major local motion problems defined below.

1) Incorrect-matched features: The feature points from current frame may be subsumed into wrong feature points of pervious frame, as indicated in the solid black circles of Fig. 1.
2) Inconsistent features: The motion of features detected from moving objects do not belong to the motion caused by the observer's hand shaking, as shown in the dotted black circles of Fig. 1.

The above local motion vectors obtained from these pairs of features may result in either sudden jitters or serious artifact effects in the reconstructed video. Therefore, a robust global motion estimation via discarding local matches is critical.

In response, this paper proposes a simple but effective video stabilization approach using shortest spanning path based motion vector clustering in order to reliably extract global motion vectors from the sets of features. Specifically, the essence of the proposed motion vector clustering is to classify local concentrations of motion vectors in feature space from corresponding feature points, that is, the different groups of motion vectors can be represented discriminatingly in which the richest group usually possesses the highest probability of containing most global motion vectors.

Experimental results via qualitative and quantitative evaluations demonstrate that the proposed video stabilization approach is able to provide a reliable estimation of global motion vectors in comparison with the other state-of-the-art approaches. In doing so, the proposed method yields a more robust and stable video viewing experience for the observers.

We organize the remainder of this paper as follows. In Section 2, we describe the genesis of our key idea, propose a shortest spanning path based video stabilization algorithm. We follow this up with qualitative and quantitative experimental results in comparisons with the other state-of-the-art video stabilization algorithms in Section 3. Finally, we conclude in Section 4.

2. Proposed video stabilization approach

In this section, we describe how the proposed method estimates global motion vectors and uses them for video stabilization. In particular, we first briefly survey the existing camera motion models for video stabilization purposes, followed by our proposed method.

2.1. Camera motion models

In general, most of video stabilization techniques to restore the unstable frame are relying on specific camera motion models, including 2-D models, 2-D affine models, 2.5-D models, and 3-D models (Yang et al., 2009). Of various camera motion models, the 2-D affine models, which possess six parameters to provide a satisfactory representation of the camera motion, are the most widely used for video stabilization while demanding less computational cost compared with both 2.5-D models and 3-D models and possessing more accuracy compared with 2-D models (Yang et al., 2009).

Hence, we adopt the 2-D affine model in proposed method to attain to satisfactory representation of the camera motion. Here, in camera coordinate system, we assume that there is a point P whose coordinates are $[x_0, y_0, z_0]^T$ and $[x_1, y_1, z_1]^T$ in the frame I_t and frame I_{t+1} before and after camera has been changed by a rotation and a translation, respectively. The relation between these two vectors of coordinates can be expressed as follows:

$$[x_1, y_1, z_1]^T = R_{3\times3} * [x_0, y_0, z_0]^T + T_{3\times1}, \tag{1}$$

where $R_{3\times3}$ and $T_{3\times1}$ represent the opposite transform of 3-D rotation and translation of camera, respectively. In addition, the coordinates of point P in frames I_t and I_{t+1} can be represented via projection as follows:

$$\left[u_0, v_0, \lambda\right]^T = \frac{\lambda}{z_0} * [x_0, y_0, z_0]^T, \tag{2}$$

$$\left[u_1, v_1, \lambda\right]^T = \frac{\lambda}{z_1} * [x_1, y_1, z_1]^T, \tag{3}$$

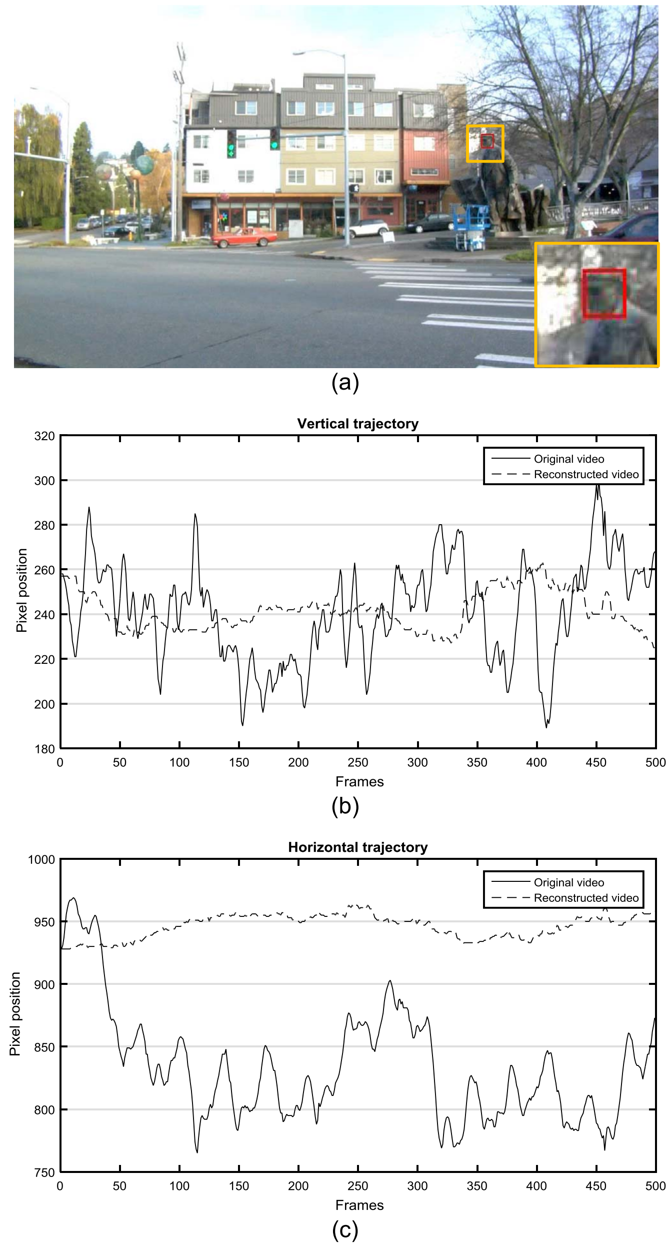

Fig. 5. Illustration of the vertical and horizontal trajectory tracking performed by the mean-shift tracking (Karami et al., 2015) using the original and reconstructed videos "Statue", respectively. (a) shows the starting point in the red bounding box for the mean-shift tracking in frame 1 of the video. (b) and (c) denote the corresponding vertical and horizontal trajectory via the mean-shift tracking from the starting point in the video, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where λ represents the distance between the camera and the captured scene in a plane. Additionally, the model can be represented by rewriting the Eqs. (1)–(3) as follows:

$$\begin{bmatrix} u_1 \\ v_1 \\ \lambda \end{bmatrix} = \frac{z_0}{z_1} * \begin{bmatrix} R_{11} R_{12} R_{13} \\ R_{21} R_{22} R_{23} \\ R_{31} R_{32} R_{33} \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ \lambda \end{bmatrix} + \frac{\lambda}{z_1} * \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}, \tag{4}$$

where the first two columns can be represented in 2-D form by

$$\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = s \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \tag{5}$$

and we define

$$s \triangleq \frac{z_0}{z_1}, \tag{6}$$

$$t_x \triangleq sR_{13}\lambda + \frac{\lambda}{z_1} T_x, \tag{7}$$

$$t_y \triangleq sR_{23}\lambda + \frac{\lambda}{z_1} T_y. \tag{8}$$

All in all, the scaling parameter s and translation parameters t_x and t_y are varying as depths change. In contrast with the depth between camera and the objects, the depth of background possesses smaller depth variation in most real scenes (Yang et al., 2009). According to this assumption, the use of this 2-D affine transform can allow us to approximate to the 3-D camera motion for video stabilization purposes.

2.2. Shortest spanning path based motion vector clustering

The main aim of the proposed method is to reliably classify

Fig. 6. Effectiveness comparison of four different video stabilization techniques applied to "Statue" video sequence. (a) The unstabilized frames. (b) Litvin et al. (2003) results. (c) Feng et al. (2013) results. (d) Yang et al. (2009) results. (e) Our results. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

global motion vectors from each frame of videos by which to reconstruct the stable frame with high robustness, limiting error variance (classifying the motion vectors of either the incorrect-matched features or inconsistent features as global motion vectors) as much as possible. By doing so, the extracted global motion vectors from corresponding feature points can be then used to determine the six parameters s, t_x, t_y, R_{11}, R_{12}, and R_{21} for stabilizing each pair of frames.

First, the feature points in this paper can be detected by using the FREAK descriptors, as suggested in Feng et al. (2013). For each feature point between two successive frames, the motion vectors can be estimated as follow:

$$M_i \in \begin{bmatrix} m_i^x \\ m_i^y \end{bmatrix}, \quad \text{where } i = 1, 2\ldots, n, \tag{9}$$

and

$$m_i^x = \|p_i^x - p_i^{x\prime}\|, \tag{10}$$

$$m_i^y = \|p_i^y - p_i^{y\prime}\|. \tag{11}$$

Note that n is the number of feature points. p_i^x and $p_i^{x\prime}$ are the horizontal positions of feature points between frame I_t and frame I_{t+1}. On the other hand, p_i^y and $p_i^{y\prime}$ are the vertical positions of feature points. Here, each motion vector can be considered as a point in a 2D subspace and the distances between these points of the 2D subspace can be calculated by using squared Euclidian norm as follows:

$$d\left(M_i, M_j\right) = \|M_i - M_j\|^2, \quad \forall\, j \neq i. \tag{12}$$

Hence, two endpoints are included into the desired path D via the shortest distance and connect them by an edge. This representative function can be expressed as follows:

$$\arg \min_{\forall j \neq i} d\left(M_i, M_j\right). \tag{13}$$

Since there are two endpoints in the desired path in the beginning, we need to decide that which path should be explored next. Hence, let us find the another two isolated points $M_a, M_b \notin D$, $\forall\, a \neq$ b that are closest to the previous two endpoints M_i, M_j by

$$d_a = \arg \min_{k \notin D} d\left(M_i, M_k\right), \tag{14}$$

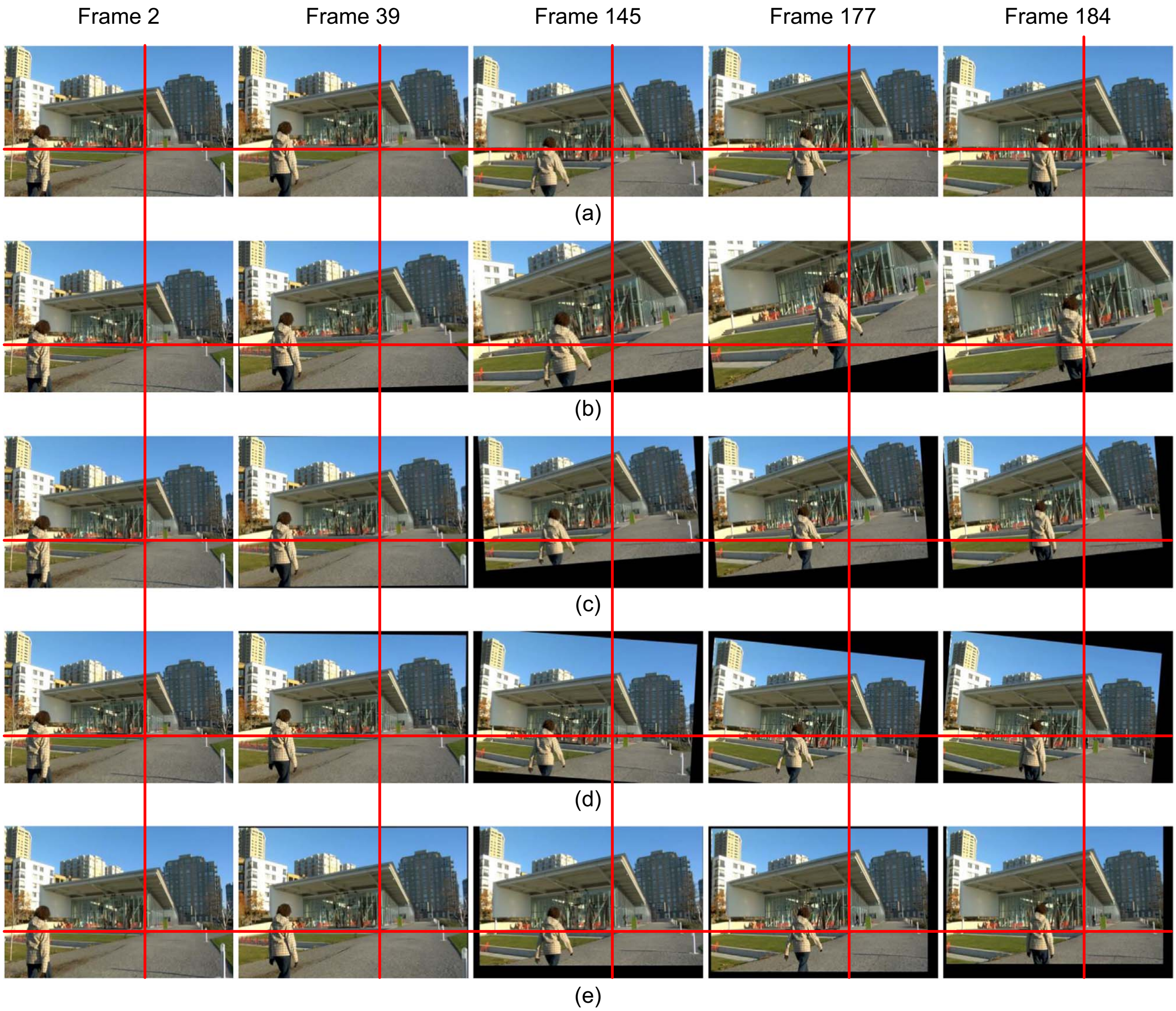

Fig. 7. Effectiveness comparison of four different video stabilization techniques applied to "Museum" video sequence. (a) The unstabilized frames. (b) Litvin et al. (2003) results. (c) Feng et al. (2013) results. (d) Yang et al. (2009) results. (e) Our results. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

$$d_b = \arg\min_{k \notin D} d\left(M_j, M_k\right). \tag{15}$$

When the distance d_b exceeds the distance d_a, the point M_a is added into desired path D and connects it by an edge with the point M_i. Otherwise, the point M_b is included into D and is connected with the point M_j. This decision rule can be expressed as

$$D \in \begin{cases} M_a, & \text{if } d_a < d_b \\ M_b, & \text{if } d_a > d_b \end{cases}. \tag{16}$$

Next, we follow the path and explore the rest of the path by using Eqs. (14)–(16) until there is no isolated points remained in the set M. In order to discriminate each cluster from the desired path D, we first calculate the mean distance d_m between all connected points in desired path D, which can be expressed as follows:

$$d_m = \frac{1}{n-1}\sum_{i=2}^{n} \|M_{i-1} - M_i\|^2, \tag{17}$$

where M_{i-1} and M_i are connected by the edge produced Eq. (16) in D. As shown in Fig. 2, when the distance between connected points exceeds the mean distance d_m plus the empirical tolerance ϵ, the edge between these two connected points is removed. Hence, the connected points will form a set of clusters C. This decision rule can be expressed as

$$C_s \begin{cases} \notin M_i, & \text{if } \|M_{i-1} - M_i\|^2 > (d_m + \epsilon) \\ \in M_i, & \text{if } \|M_{i-1} - M_i\|^2 < (d_m + \epsilon) \end{cases}, \\ s.\,t.\;\; \forall\, i \neq 1 \wedge M_{i-1}, M_i \in D, \tag{18}$$

where C_s is the set of clusters, $s = 1, 2, \ldots, H$, and H is the number of clusters. As mentioned above, the flowchart of the proposed method is shown in Fig. 3.

After the points are discriminated into each cluster, the feature points belonging to the richest cluster are selected since it possesses the highest probability of containing the global motion vectors. Fig. 4 (a) and (b) show that the comparison with and without using the proposed method for feature classification, respectively. Thanks to the use of the proposed method, the motion vectors which are extracted from both the incorrect-matched and inconsistent feature points are eliminated.

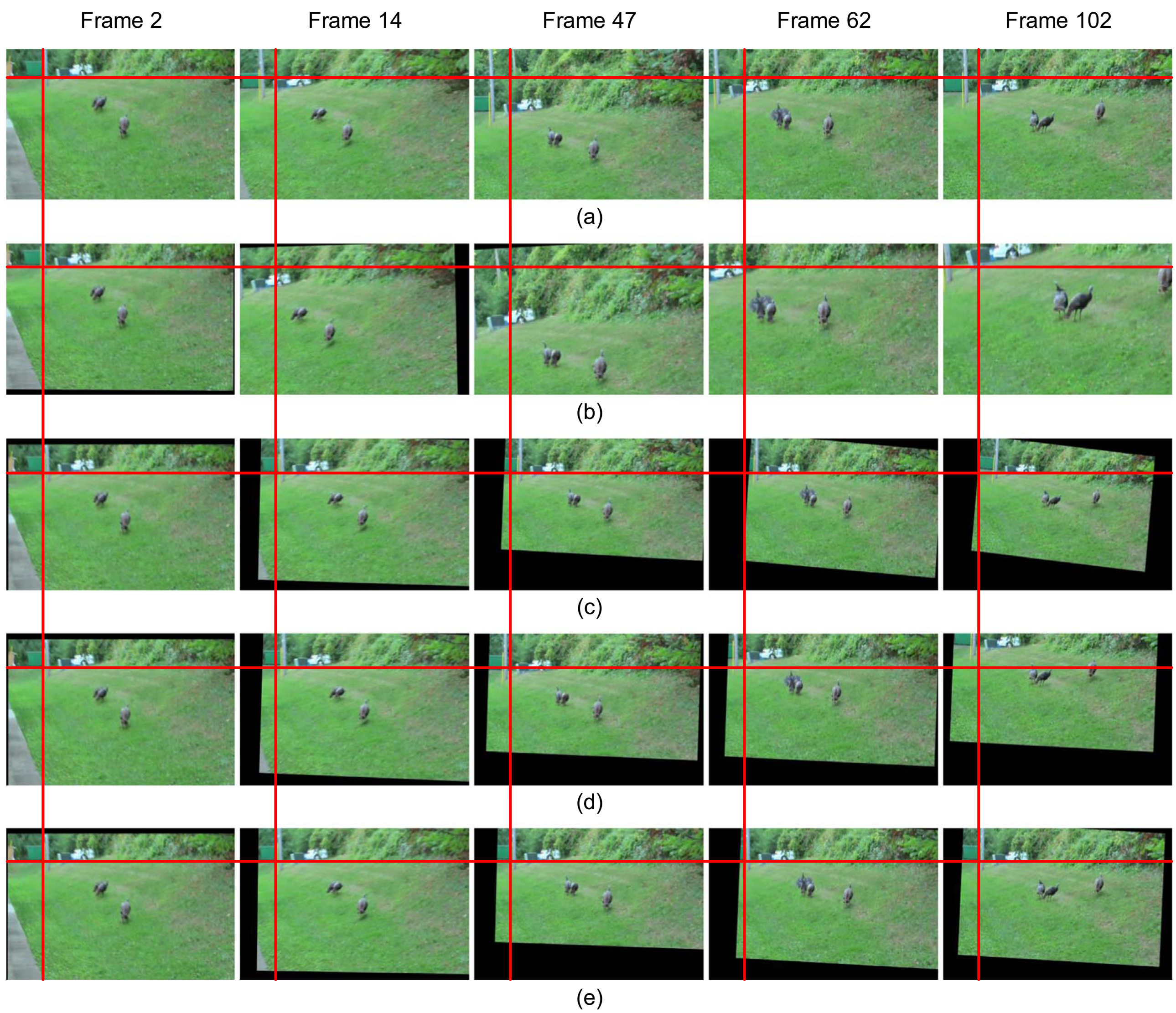

Fig. 8. Effectiveness comparison of four different video stabilization techniques applied to "Duck" video sequence. (a) The unstabilized frames. (b) Litvin et al. (2003) results. (c) Feng et al. (2013) results. (d) Yang et al. (2009) results. (e) Our results. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Additionally, we regard these remained motion vectors belonging to unintended camera motions as global motion vectors by which to calculate the 2-D affine transform parameters for each pair of frames and recover the unstable frames. Therefore, the proposed method is able to provide a reliable estimation with low error variance for the global motion vectors. As indicated via the vertical and horizontal trajectory tracking performed by the mean-shift tracking (Karami et al., 2015) in Fig. 5, the proposed method is capable of achieving more robust video stabilization without unexpected jitter or undesired artifact effect.

3. Experimental results

In this section, the effectiveness of the proposed method is evaluated via qualitative and quantitative measurements and compared to the three other state-of-the-art methods, including methods of Feng et al. (2013), Yang et al. (2009), and Litvin et al. (2003). To this end, the methods of Yang et al. (2009) and Litvin et al. are modified to employ the FREAK descriptors by which to detect feature points from two successive frames. Moreover, five video sequences[1,2] captured in realistic scenes with unexpected image motion caused by the observer's hand shaking were used for testing, named "Statue", "Museum", "Duck", "Lab", and "Sidewalk".

3.1. Qualitative measurement

The goal of this part is to conduct a visual assessment of the reconstructed frames produced by each compared method. Figs. 6–10 show five examples that demonstrate the respective effectiveness of the proposed method and the other compared methods. The red lines in Figs. 6–10 (a)–(e) indicate the horizontal and vertical planes for visualization of the reconstructed results, respectively.

As can be observed in Figs. 6–10(a), each compared method employed the FREAK descriptors to stabilize these video sequences. First of all, Litvin et al. (2003) method minimized the p-norm cost function based on multiscale implementation in order

[1] http://web.cecs.pdx.edu/~fliu/project/subspace_stabilization/.
[2] http://changedetection.net/.

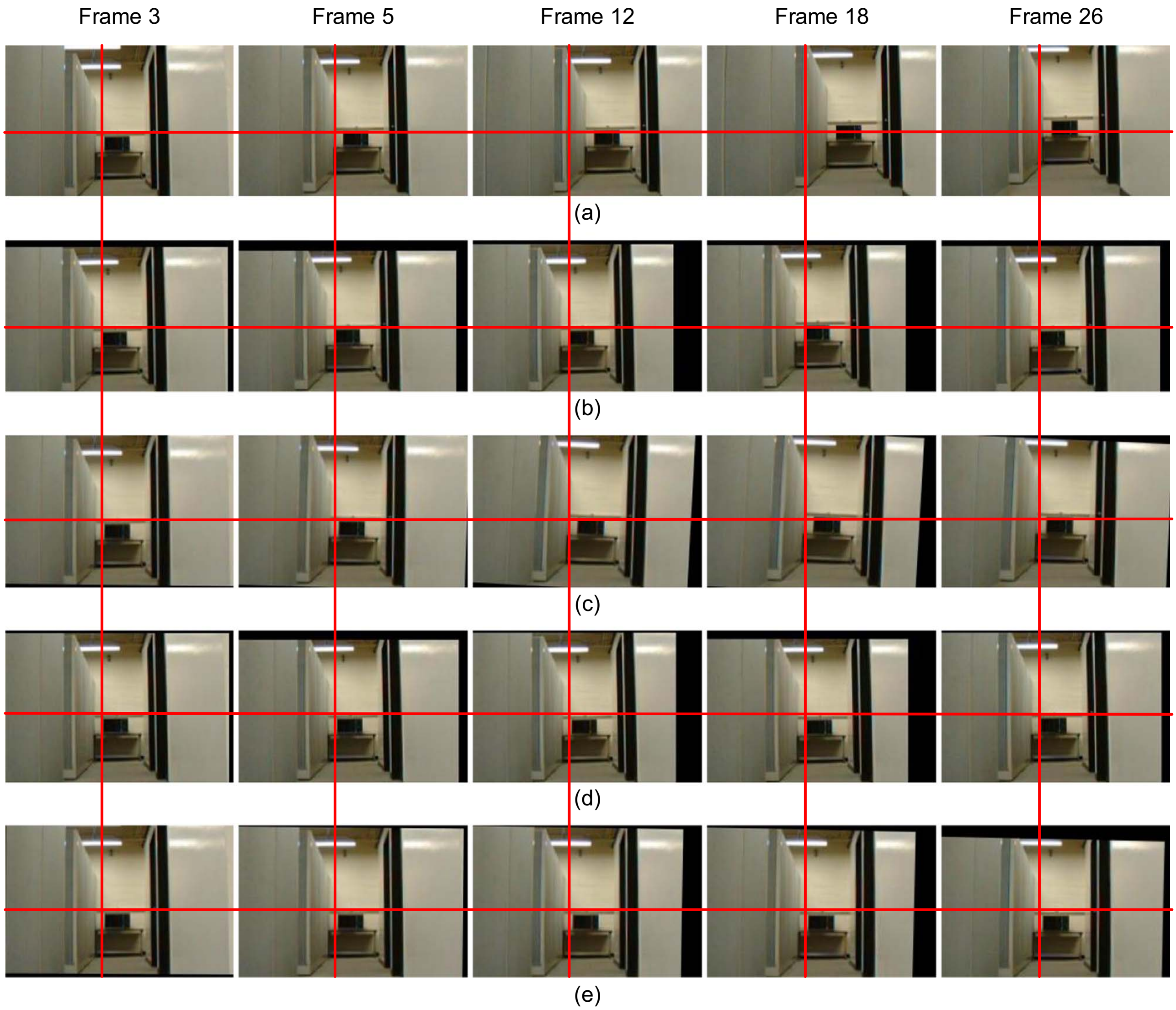

Fig. 9. Effectiveness comparison of four different video stabilization techniques applied to "Lab" video sequence. (a) The unstabilized frames. (b) Litvin et al. (2003) results. (c) Feng et al. (2013) results. (d) Yang et al. (2009) results. (e) Our results. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

to estimate the global motion vectors. However, this resulted in the estimated global motion vectors based on different scales were unstable between each pair of frames. Hence, the method of Litvin et al. (2003) inevitably failed to stabilize videos, as indicated in Figs. 6–10(b). Feng et al. (2013) method and Yang et al. (2009) method utilized the MSAC technique and the particle filters to attain the global motion vectors between successive frames, respectively. However, since the uses of the MSAC technique and the particle filters were not able to accurately separate the global motion vectors from the all detected motion vectors, the frames reconstructed by using the methods of Feng et al. (2013) and Yang et al. (2009) were unstabilized, as indicated by the red lines in Figs. 6–10(c) and (d).

In contrast to these methods, the proposed method is able to reconstruct each frame with more stable effectiveness, as can be seen in Figs. 6–10(e). This is due to the ability of the proposed method to accurately classify the detected motion vectors by using the shortest spanning path clustering into several clusters between successive frames, whereupon the proposed method employed the richest cluster which possessed the highest probability of containing most of global motion vectors to stabilize each frame. Consequently, the superior stabilized effectiveness can be achieved by the proposed method.

3.2. Quantitative measurement

In this part, we presented the respective effectiveness of each compared method as determined via quantitative measurement for each test video sequence. To this end, the MSE, PSNR, and SSIM metrics were used to demonstrate the effectiveness of video stabilization in each pair of successive frames (Huang et al., 2013; Morimoto and Chellappa, 1998). Tables 1–3 summarized the comparisons between the average MSE, PSNR, and SSIM rates produced through each compared method on video sequences "Statue", "Museum", "Duck", "Lab", and "Sidewalk", respectively. Lower MSE values signify superior stabilized efficacy. In contrast, lower PSNR and SSIM values signify inferior stabilized efficacy.

As can be observed in Tables 1–3, the Litvin et al. (2003) method was not able to stabilize videos and made these videos even more unstable. Although the reconstruction results show improvement when adopting the methods of Yang et al. (2009) and Feng et al. (2013), the results still have room for stabilized efficacy. On the other hand, the proposed method attained the best stabilized effectiveness in comparison with the other compared

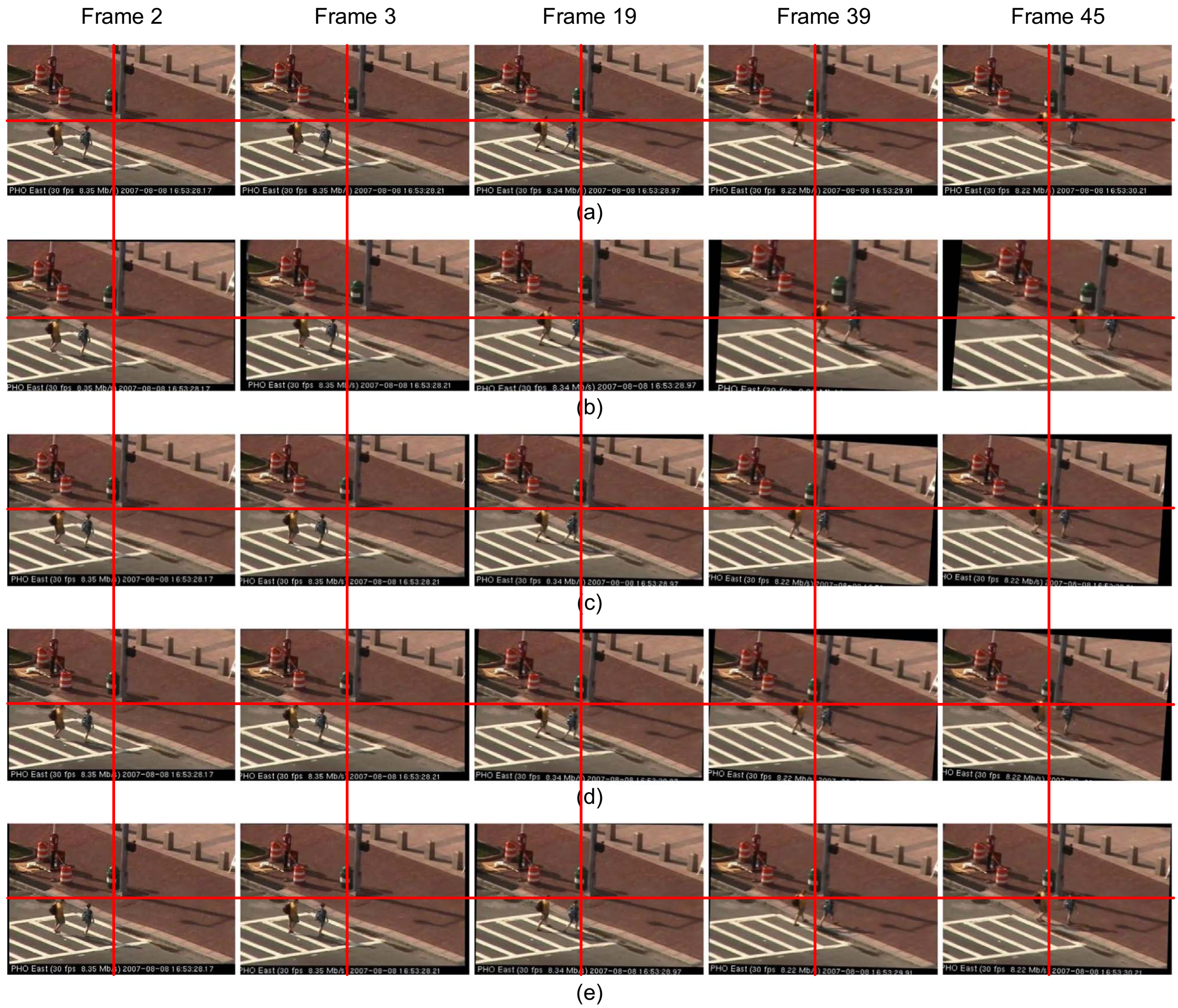

Fig. 10. Effectiveness comparison of four different video stabilization techniques applied to "Sidewalk" video sequence. (a) The unstabilized frames. (b) Litvin et al. (2003) results. (c) Feng et al. (2013) results. (d) Yang et al. (2009) results. (e) Our results. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 1
MSE comparison of stabilized effectiveness for each video sequence.

Methods	Statue	Museum	Duck	Lab	Sidewalk
Unstable video	868.63	679.19	381.51	985.10	1155.63
Litvin et al. (2003)	1672.53	1077.79	605.27	855.51	2355.39
Feng et al. (2013)	326.82	256.85	421.69	719.37	742.41
Yang et al. (2009)	388.96	266.46	404.28	761.90	733.50
Ours	**307.72**	**207.29**	**285.20**	**706.20**	**710.68**

Table 2
PSNR comparison of stabilized effectiveness for each video sequence.

Methods	Statue	Museum	Duck	Lab	Sidewalk
Unstable video	21.84	20.11	22.71	18.45	18.97
Litvin et al. (2003)	16.01	18.38	20.93	19.12	14.74
Feng et al. (2013)	25.84	24.60	23.08	20.51	19.79
Yang et al. (2009)	22.58	24.07	23.34	19.49	19.81
Ours	**26.04**	**25.27**	**24.45**	**20.59**	**19.88**

Table 3
SSIM comparison of stabilized effectiveness for each video sequence.

Methods	Statue	Museum	Duck	Lab	Sidewalk
Unstable video	0.6047	0.6279	0.5268	0.6526	0.6001
Litvin et al. (2003)	0.5196	0.5847	0.6148	0.8304	0.4215
Feng et al. (2013)	0.9404	0.8733	0.8725	0.8401	0.8513
Yang et al. (2009)	0.9290	0.8749	0.7709	0.8607	0.8443
Ours	**0.9505**	**0.9032**	**0.8845**	**0.8609**	**0.8567**

methods according to this sophisticated quantitative measurement, meaning that the proposed method is capable of providing superior and reliable video stabilization for all test videos.

4. Conclusions

In this paper, we presented a shortest spanning path based video stabilization method by which to reliably remove the unexpected image motion caused by the observer's hand shaking. To this end, feature points detected by the FREAK descriptors from each pair of frames were used to calculate the global motion vectors through the shortest spanning path clustering algorithm, after which the global motion vectors were used to reliably stabilize the successive frames. The stabilized effectiveness of the proposed method has been demonstrated through both qualitative and quantitative evaluations conducted via the use of five videos in which the proposed method achieves superior stabilized effectiveness in comparison with the other state-of-the-art methods.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan, and Russian Foundation for Basic Research, Russia, under Grant nos. MOST 105-2923-E-027-001-MY3, MOST 103-2221-E-027-031-MY2, MOST 103-2221-E-027-030-MY2, MOST 103-2923-E-002-011-MY3, MOST 104-2221-E-027-020, MOST 105-2218-E-155-003, MOST 104-3115-E-155-002, MOST 105-2811-E-027-001, RFBR 16-57-52042, and RFBR 14-07-00527.

References

Alahi, A., Ortiz, R., Vandergheynst, P., Jun 2012. FREAK: Fast retina keypoint. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 510–517.
Battiato, S., Gallo, G., Puglisi, G., Scellato, S., Sept. 2007. SIFT features tracking for video stabilization. In: Proceedings of the International Conference on Image Analysis and Processing. pp. 825–830.
Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110 (Jun. (3)), 346–359.
Chang, H.C., Lai, S.H., Lu, K.R., Jun 2004. A Robust and Efficient Video Stabilization Algorithm. Proc. Int. Conf. on Multimedia and Expo. pp. 29–32.
Feng, K., Yonghua, H., Huaxiong, Z., Dec 2013. Video Stabilization Based on Multiscale Local Color Invariants. In: Proceedings of the International Conference on Networking and Distributed Computing. pp. 65–69.
Huang, S.C., Cheng, F.C., Chiu, Y.S., 2013. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE Trans. Image Process. 22 (Mar.), 1032–1041.
Kang, S.J., Wang, T.S., Kim, D.H., Morales, A., Ko, S.J., Feb 2012. Video stabilization based on motion segmentation. In: Proceedings of the IEEE International Conference on Consumer Electronics. pp. 416–417.
Karami, A.H., Hasanzadeh, M., Kasaei, S., 2015. Online adaptive motion model-based target tracking using local search algorithm. Eng. Appl. Artif. Intell. 37 (Jan.), 307–318.
Kherallah, M., Bouri, F., Alimi, A.M., 2009. On-line Arabic handwriting recognition system based on visual encoding and genetic algorithm. Eng. Appl. Artif. Intell. 22 (1), 153–170.
Ko, S., Lee, S., Jeon, S., Kang, E., 1999. Fast digital image stabilizer based on gray-coded bit-plane matching. IEEE Trans. Consum. Electron. 45 (Aug. (3)), 598–603.
Kwon, O., Shin, J., Paik, J., Jan. 2006. Edge based Adaptive Kalman Filtering for Real-Time Video Stabilization. in: Proc. Int. Conf. Consum. Electron. pp. 75–76.
Leu, J.S., Lin, W.H., Tzeng, H.J., Chen, C.F., Lin, M.S., 2012. Adaptive frame synchronization for surveillance system across a heterogeneous network. Eng. Appl. Artif. Intell. 25 (7), 1349–1354.
Litvin, A., Konrad, J., Karl, W., Jan 2003. Probabilistic video stabilization using kalman filtering and mosaicking. In: Proceedings of the Image Video Communication IS&T/SPIE Symposium on Electronics Imaging. pp. 663–674.
Lowe, D.G., Sept 1999. Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157.
Matsushita, Y., Ofek, E., Ge, W., Tang, X., Shum, H.-Y., 2006. Full-frame video stabilization with motion inpainting. IEEE Trans. Pattern Anal. Mach. Intell. 28 (7), 1150–1163.
Morimoto, C., Chellappa, R., 1998. Evaluation of image stabilization algorithms. In: Proceedings of the IEEE International Conference. Acoust., Speech, Signal Process. 5. pp. 2789–2792.
Niskanen, M., Silven, O., Tico, M., Jul. 2006. Video Stabilization Performance Assessment. In: Proceedings of the International Conference on Multimedia and Expo. pp. 405–408.
Pandian, S.I., Bala, G.J., Anitha, J., 2013. A pattern based PSO approach for block matching in motion estimation. Eng. Appl. Artif. Intell. 26 (8), 1811–1817.
Pinto, B., Anurenjan, P.R., Feb. 2011. Video stabilization using speeded up robust features. Proc. IEEE Int. Conf. Commun. Signal Process. pp. 527–531.
Puglisi, G., Battiato, S., 2011. A robust image alignment algorithm for video stabilization purposes. IEEE Trans. Circuits Syst. Video Technol. 21 (10), 1390–1400.
Qu, H., Song, L., Xue, G., Nov. 2013. Shaking video synthesis for video stabilization performance assessment. In: Proceedings of the Visual Communications and Image Processing. pp. 1–6.
Shen, Y., Guturo, P., Damarla, T., Buckles, B.P., Namuduri, K.R., 2009. Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework. IEEE Trans. Consum. Electron. 55 (3), 1714–1721.
Shi, J., Tomasi, C., Jun. 1994. Good features to track. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition. pp. 593–600.
Yang, J., Schonfeld, D., Mohamed, M., 2009. Robust video stabilization based on particle filter tracking of projected camera motion. IEEE Trans. Circuits Syst. Video Technol. 19 (7), 945–954.