



# Motion energy image for evaluation of video stabilization

Marcos Roberto e Souza<sup>1</sup> · Helio Pedrini<sup>1</sup>

Published online: 27 June 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Large volumes of video content have been generated through the development of compact and portable cameras. Examples of applications that have been benefited from such growth of multimedia data include business conferencing, telemedicine, surveillance and security, entertainment, distance learning and robotics. Video stabilization is the process of detecting and removing undesired motion or instabilities from a video stream caused during the acquisition stage when handling the camera. In this work, we introduce and analyze a novel visual representation based on motion energy image for qualitative evaluation of video stabilization approaches. Experiments conducted on different video sequences are performed to demonstrate the effectiveness of the visual representation as qualitative measure for evaluating video stability.

**Keywords** Motion energy image · Video stabilization · Handheld camera · Qualitative evaluation

## 1 Introduction

Several mobile devices capable of capturing videos in various circumstances have become popular in recent years. The handling of these devices is typically not controlled and may include unwanted oscillations, degrading video quality.

Video stabilization [2,6,13,17,19–21,25,27,32] aims to compensate for undesired motion of the camera during video acquisition. Efficient methods for stabilizing videos are important to improve their quality according to human perception, as well as facilitate other tasks such as indexing, search and content retrieval [11,12,16].

Techniques and criteria for evaluation of video stabilization must be well established to leverage the state-of-the-art in the field, such that approaches can be improved and compared in an appropriate manner.

Quantitative techniques for the evaluation of video stabilization available in the literature are, in some cases, incompatible with visual perception. In addition, the methods used to evaluate and report the results subjectively are little explored. As main contribution of this work, we introduce and analyze the motion energy image (MEI) for the subjective and objective evaluation of video stabilization methods. One of the reasons why we proposed to apply such method

to assess the final results is that it would be very complex to incorporate it as part of the video stabilization pipeline itself. On the other hand, the evaluation is not biased by the technique used to stabilize the video.

The proposed method considers that the video stabilization quality can be estimated by the amount of motion present in the video. Both low-frequency motion and high-frequency motion present in the video are considered, since they may contain instabilities [15]. Object movement is also taken into account; however, this does not represent a problem since the proposed method is used to compare results for a same video sequence (before and after the stabilization process) and not to compare different video sequences. Experimental results show that our method is efficient to evaluate stabilization, differentiating stable from unstable videos. The assessment of the results is more coherent when compared to the metrics used in the literature.

This paper is organized as follows. Relevant concepts and related work are briefly described in Sect. 2. The use of the motion energy image for subjective evaluation of video stabilization is presented in Sect. 3. Experimental results are presented and discussed in Sect. 4. Final remarks and directions for future work are described in Sect. 5.

✉ Helio Pedrini  
helio@ic.unicamp.br

<sup>1</sup> Institute of Computing, University of Campinas, Campinas, SP 13083-852, Brazil

## 2 Background

Video stabilization systems of different categories have been proposed in the literature, where the most common are mechanical stabilization, optical stabilization and digital stabilization. Mechanical stabilization uses sensors to detect and compensate for camera shake. Optical stabilization consists of a mechanism to compensate for angular and translational motion, stabilizing the image before it is recorded on the sensor. Digital stabilization is implemented in software without the use of special devices.

In the context of image and video processing, the evaluation process can be classified as either objective or subjective. Evaluation is objective when measured by means of quantitative metrics applied between two images or videos, whereas evaluation is subjective when it is performed by human observers [14]. In both cases, it is typically desired to evaluate stabilization using criteria based on the perception of the human visual system.

### 2.1 Objective evaluation

Criteria for measuring the amount and nature of the displacement have been proposed to evaluate the quality of video stabilization objectively [28]. Unintentional motion is decomposed into divergence and jitter through low-pass and high-pass filters, respectively. The amount of jitter from the stabilized and original video is compared, as well as the divergence is verified, which indicates the amount of expected displacement. As an overall assessment, the blurring caused by the stabilization process is also considered.

Most of the approaches available in the literature adopt the Interframe Transformation Fidelity (ITF) metric [3,8,10,29,35], which can be expressed as the average of peak signal-to-noise ratio (PSNR) for each pair of frames in the video. Some recent techniques consider the Structural Similarity (SSIM) index [36] as an alternative to PSNR [8].

Liu et al. [23] employed the amount of energy present in the low-frequency portion of the estimated 2D motion as a stability metric. Rates of frame cropping and distortion were also considered to assess stabilization more generally.

The synthesis of unstable videos from stable videos was proposed to evaluate the stabilization process [30] in order to have the ground-truth of the stable videos. The methods developed in their work were evaluated according to two aspects: (i) the distance between the stabilized frame and the reference frame and (ii) the average of the SSIM between each pair of consecutive frames.

Due to the weaknesses of the ITF metric in motion videos, a measure based on the variation of the intersection of angles between the global motion vectors calculated from the SIFT keypoints [24] was developed to evaluate the video stabilization process [7]. In fixed-camera videos, the ITF metric was

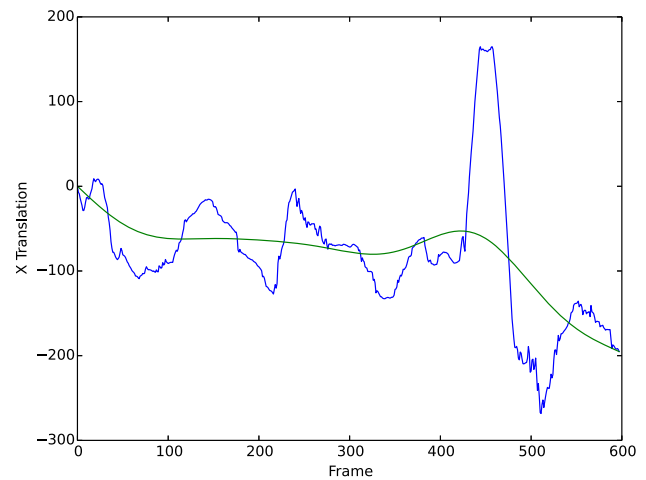


Fig. 1 Trajectory of horizontal translation

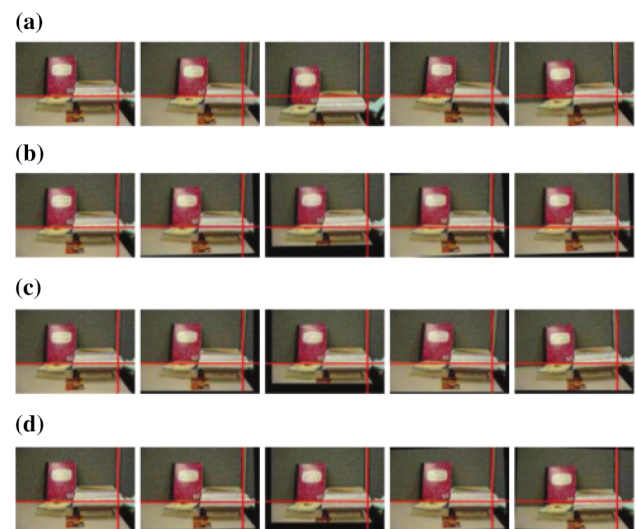


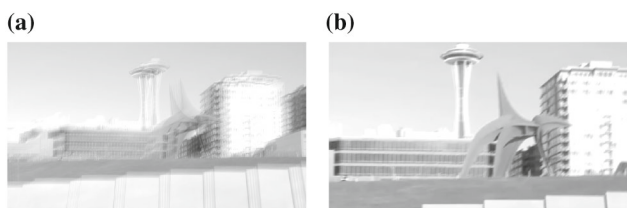
Fig. 2 Frame sequence of a video. **a** Original video, **b–d** different versions of the stabilized video. Extracted from [37]

considered, however, only for overlapping the frame background instead of the entire frame.

### 2.2 Subjective evaluation

Many approaches have briefly analyzed the trajectories made by the camera and the trajectories of the stabilized video [5,9,22,26,31]. These trajectories are usually related to the different factors that compose the estimated 2D motion. Some methods present, for example, the path of the camera through transformations, for instance, translations and rotations. Figure 1 shows an example of horizontal translation path estimated from the original (green) and smoothed (blue) videos.

From the trajectory, we can observe when a motion occurs and its intensity in the original video, as well as such motion



**Fig. 3** Mean gray-level frames for the first ten frames. **a** Original video; **b** stabilized video



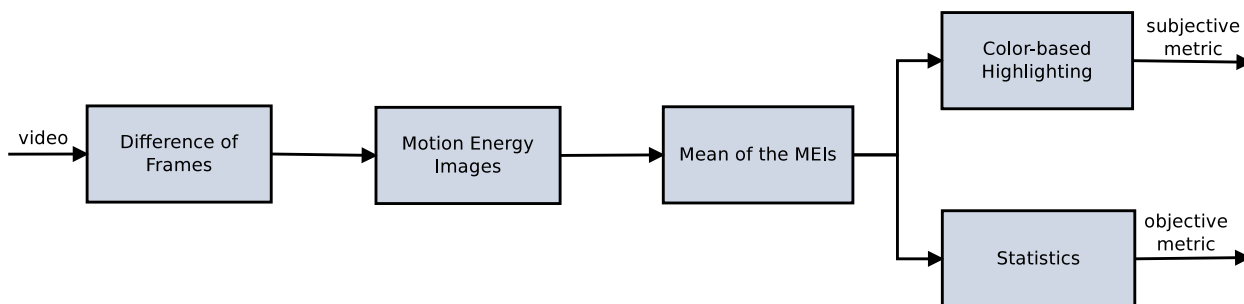
**Fig. 4** Histograms of motion in the HSV color space. Extracted from [33]

after its smoothing. This type of visualization can be very useful to analyze the behavior of the smoothing step of the motion employed in a certain method. However, its result depends on the technique used in the motion estimation, so that the trajectory does not accurately represent the video motion. Thus, the trajectory may be neither a good alternative for the evaluation of stabilization quality nor an adequate visualization for videos with spatially distinct motion.

Other approaches show frame sequences, usually superimposed by horizontal and vertical lines [5,8,10,22,26,34,37]. Thus, it is possible to check the alignment of a small set of consecutive frames. Figure 2 illustrates an example of this view, where objects intercepted by lines must be more aligned in the stabilized video.

From the sequence of frames, it is possible to analyze the displacement of each frame, in addition to the amount of pixels lost due to the transformation applied to each frame. However, this technique becomes impractical when a large number of frames are considered, making it unfeasible to analyze the entire video.

There are also some approaches that summarize the video with a single image through the mean gray-level frame [18,38], as shown in Fig. 3. Sharper images are expected for more stable videos. From this view, it is possible to check if the



**Fig. 5** Main steps of the proposed methodology



**Fig. 6** Pseudocolor transformation applied to the images of the average MEIs

video has motion; however, it is difficult to determine the nature of the motion present in the video.

In a broader context, the visualization of videos is concerned with the creation of a new visual representation, obtained from an input video, capable of indicating its characteristics and important events [4]. Video visualization techniques can generate different types of output data, such as another video, a collection of images, a single image, among others. Borgo et al. [4] reported a review of several video visualization techniques proposed over the last years.

In order to help users find scenes with specific motion characteristics in the context of video browsing, the visualization of motion histograms in the hue-saturation-value (HSV) color space [33] was proposed. The motion histograms were obtained with the motion vectors contained in H.264 / AVC. Figure 4 shows an example of such visualization, where each frame of the video is represented by a vertical line, such that the motion direction is mapped into different colors and the motion intensity mapped into brightness values. The disadvantage of this technique is the presence of noise in the motion vectors, introduced by the motion estimation algorithm [33].

### 3 Average motion energy image for subjective evaluation

The motion energy image (MEI) is a binary image that represents the presence of video motion in a given region. This occurrence is determined by the difference in the gray levels of the video frames. White pixels denote the presence of motion, whereas black pixels denote its absence [1]. In conjunction with the motion history image (MHI), MEI is

**Table 1** Video sequences from the first dataset

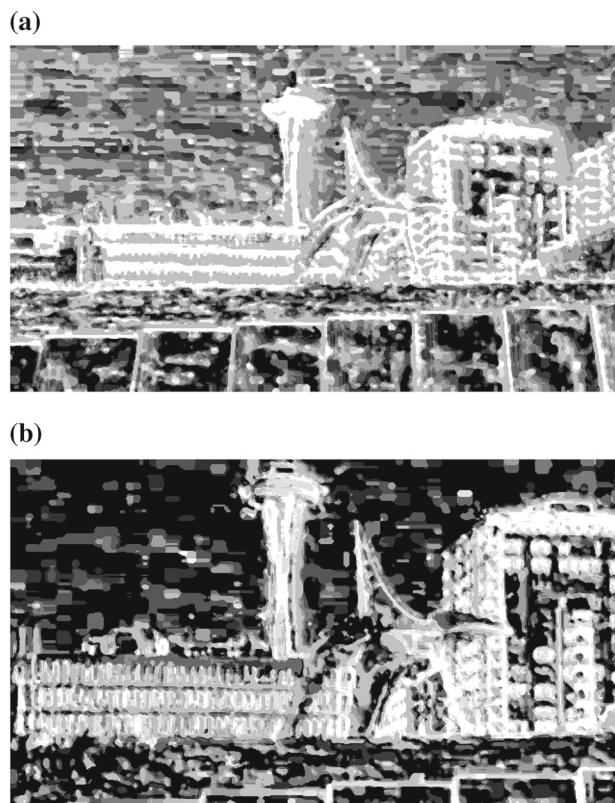
#	Video	Source
1	gleicher1	GaTech VideoStab
2	gleicher2	GaTech VideoStab
3	gleicher3	GaTech VideoStab
4	gleicher4	GaTech VideoStab
5	greyson_chance	GaTech VideoStab
6	hippo	nghiaho.com/uploads/hippo.mp4
7	If_juggle	GaTech VideoStab
8	new_gleicher	GaTech VideoStab
9	sam_1	GaTech VideoStab
10	sam and cocoa	youtu.be/627MqC6E5Yo
11	sany0025	GaTech VideoStab
12	shake_pgh_1	GaTech VideoStab
13	shaky_car	Matlab
14	yuna_long	GaTech VideoStab

**Table 2** Categories and amount of videos present in the second dataset, proposed by Liu et al. [23]

Category	# Videos
Crowd	22
Parallax	17
Quick rotation	28
Regular	22
Running	21
Zooming	29
Total	139

generally used in the context of human action recognition in videos [1].

In this work, we consider the average of the motion energy images obtained throughout the video to assess the amount of motion and to characterize its stability. Figure 5 presents the main stages of our methodology.



**Fig. 9** MEI for video #4 with  $i = 2$ . **a** Original video, **b** stabilized video

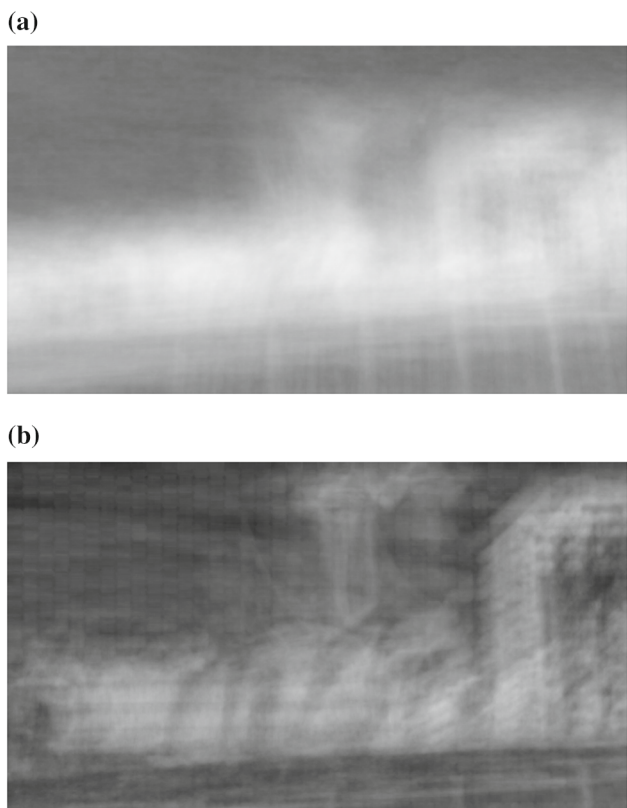
For each video frame  $i$ , the difference of the gray levels of each pixel is calculated. This is done by considering the preprocessed frames through a Gaussian filter with kernel experimentally set as  $\sigma = 5$ , which is applied to smooth the frames, so that the difference is calculated without disregarding unnecessary details. In this step, a binary image is obtained, where 1 is assigned to the pixel with difference greater than a certain threshold, and 0 otherwise. This calculation can be seen as a sub-step of the MEI construction,



**Fig. 7** Difference images for unstable video #4 with  $i = 2$ . **a**  $j = 3$ , **b**  $j = 4$ , **c**  $j = 5$ , **d**  $j = 6$ , **e**  $j = 7$



**Fig. 8** Difference images for video #4 after stabilization with  $i = 2$ . **a**  $j = 3$ , **b**  $j = 4$ , **c**  $j = 5$ , **d**  $j = 6$ , **e**  $j = 7$



**Fig. 10** Average image of the MEIs for video #4. **a** Original video, **b** stabilized video

expressed as

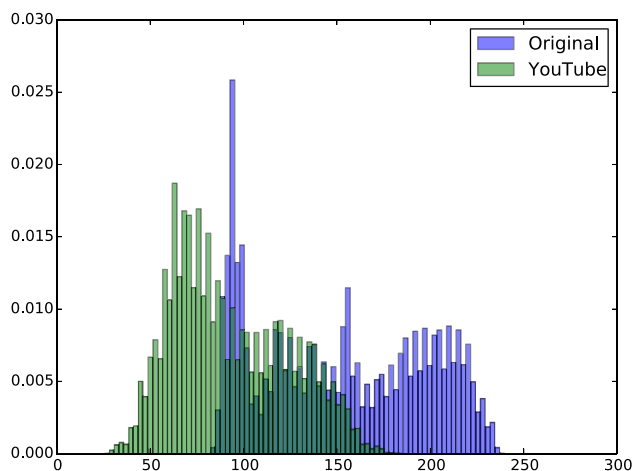
$$Diff_{i,j}(x, y) = \begin{cases} 1 & \text{if } \text{med}(|f_i(x, y) - f_j(x, y)|) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where  $(x, y)$  denotes a given pixel and  $f$  is the already smoothed frame. In turn,  $i$  and  $j$  correspond to the  $i$ -th and  $j$ -th frame indices, respectively.  $T$  corresponds to the threshold, experimentally chosen as 10. Finally,  $\text{med}(\cdot)$  is a median filter with kernel of size 5, applied to decrease the discontinuities of the differences.

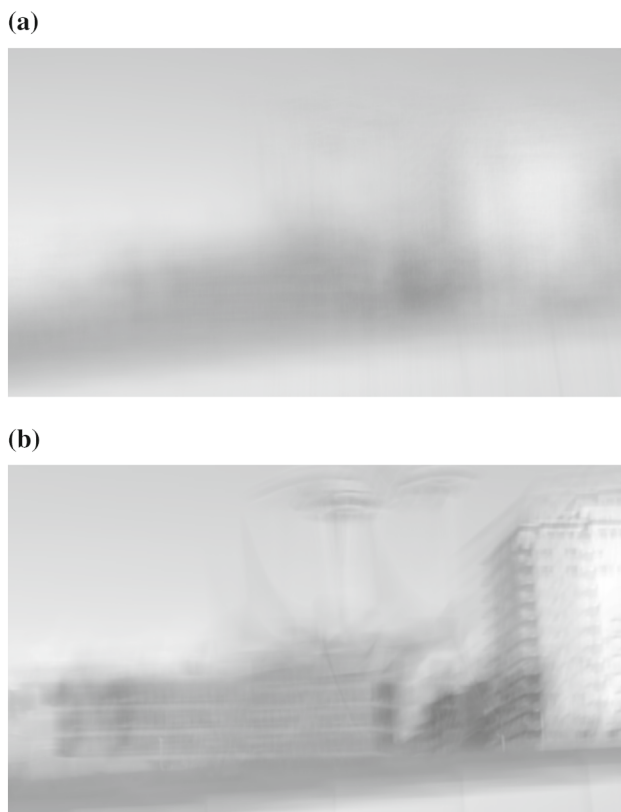
We consider an MEI for each frame  $i$ , which is obtained through the differences of the frames within a sliding window of size  $N$ , centered in  $i$ . The MEI calculation can be expressed as

$$MEI_i = \frac{\sum_{j \in \Omega_i, j \neq i} G(|i - j|, \sigma) Diff_{i,j}}{\sum_{j \in \Omega_i, j \neq i} G(|i - j|, \sigma)} \tag{2}$$

where  $G(\cdot)$  is a Gaussian function that assigns larger weights to the differences of the nearest frames.  $\Omega_i$  is the neighborhood of  $i$  determined by the sliding window.



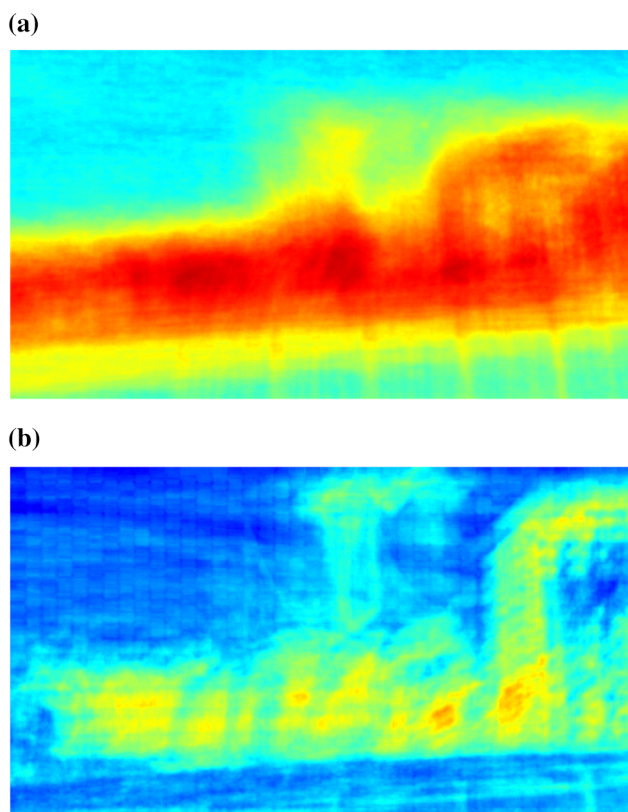
**Fig. 11** Histogram of average image of MEIs for video #4



**Fig. 12** Average image of the colored MEIs for video #4. **a** Original video, **b** stabilized video

In contrast to the MEI calculation typically performed in the literature, we consider the differences from the central window frame. This is done so that motion that occurs more gradually can be captured by MEI.

The window size  $N$  is based on the number of frames per second (FPS), in order to always consider the same time



**Fig. 13** Average grayscale image for video #4. **a** Original video, **b** stabilized video

interval, expressed as

$$N = \frac{\text{FPS}}{n} \quad (3)$$

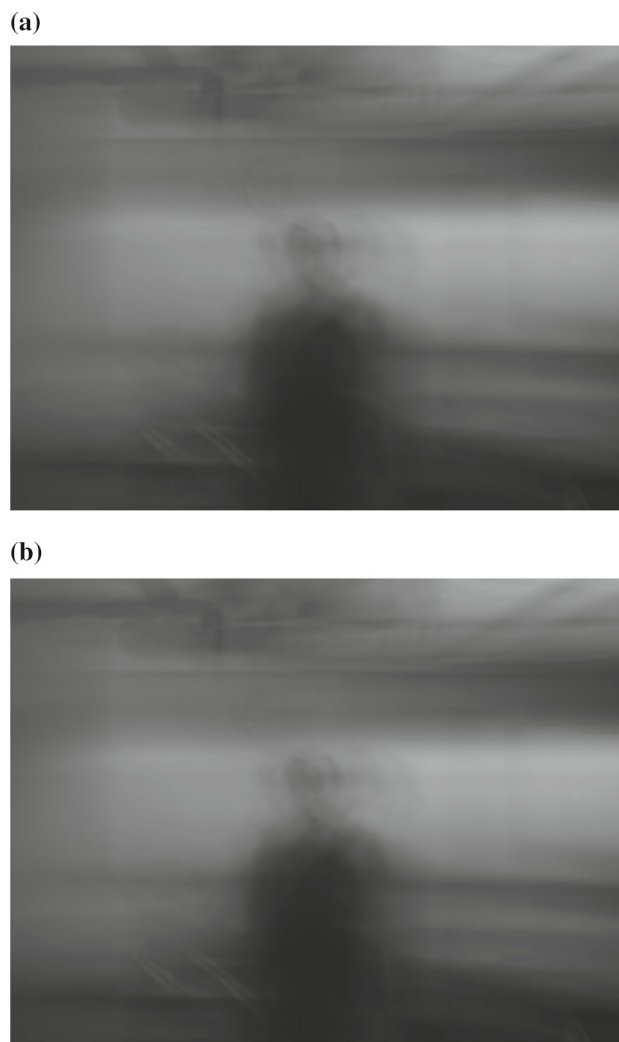
where  $n = 5$  is empirically adopted in our work.

The use of a Gaussian function to provide larger weights for the frames closer to the central frame is premised on the fact that oscillations present in unstable videos usually occur more suddenly than a desired motion.

By taking the MEI of each frame, the average image of the MEIs is calculated, where each pixel  $(x, y)$  is taken as the arithmetic mean of the pixels  $(x, y)$  of all MEIs of the video. Thus, from the gray-level image obtained, it is possible to verify the amount of motion present in the video, its location and spatial distribution in the frames.

The human visual system can distinguish thousands of colors, however, only a few tens of shades of gray. Thus, a pseudocolor transformation is applied, so that high gray-level intensity values are mapped to red, whereas lower intensities to blue. Figure 6 shows the color mapping used. A more stable video is expected to have less motion, and therefore, a view with colors that are closer to blue than an unstable video is obtained.

In addition to the visualization, we extracted statistical measurements from the gray-level image in order to obtain



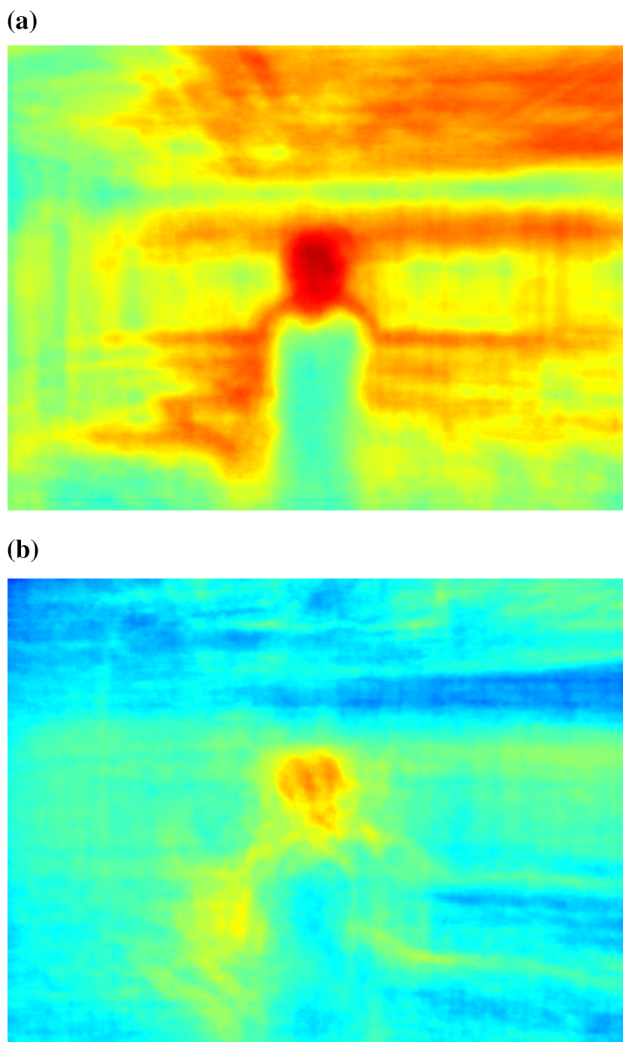
**Fig. 14** Average grayscale image for video #7. **a** Original video, **b** stabilized video

an objective metric that characterizes the average amount of motion (AAM) present in the video and that can be used to determine the quality of the stabilization process. For this, we consider the normalized average of the gray-level intensities, which can be expressed as

$$\text{AAM} = \frac{\sum_x^W \sum_y^H I(x, y)}{WHL_{\max}} \quad (4)$$

where  $W$  and  $H$  correspond to the width and height of the image, respectively, whereas  $L_{\max}$  is the maximum intensity that a pixel can assume.

The AAM value is normalized between 0 and 1. Higher values indicate a greater amount of motion. Typically, a more stable video should generate a lower AAM value than its unstable version. For visualization purpose, we used the AAM to compare videos before and after the stabilization



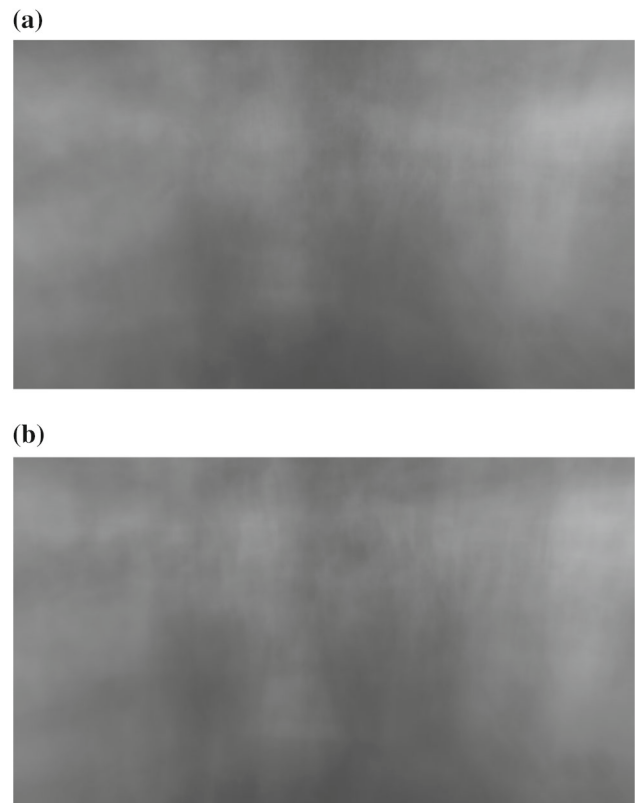
**Fig. 15** Average image of the colored MEIs for video #7. **a** Original video, **b** stabilized video

process. Therefore, we need not be concerned with the interference of moving objects, since this will occur in both videos.

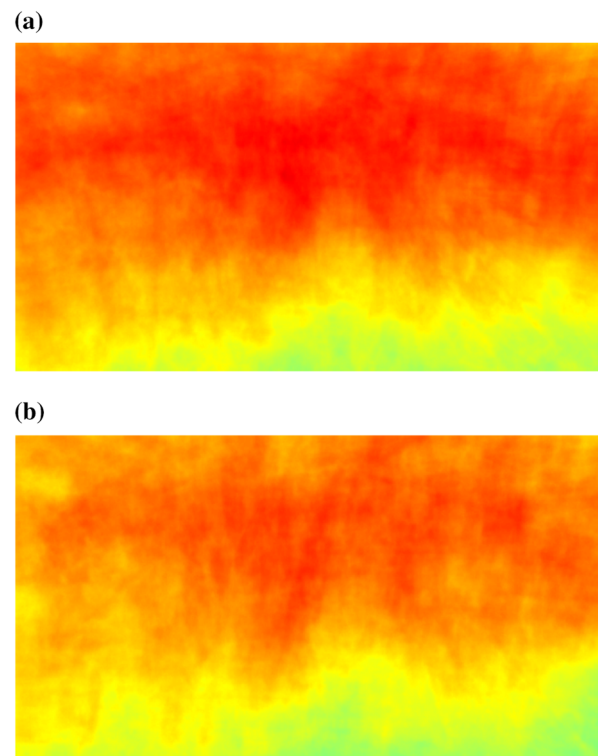
## 4 Results

This section presents the results obtained from our experiments. Sects. 4.1 and 4.2 describe the results obtained with the subjective visualization and the objective metric, respectively.

Two datasets were used in our experiments. The first one is composed of fourteen videos, where eleven videos were extracted from the GaTech VideoStab [15] and the others collected separately. Table 1 reports a summary of the first database with videos in alphabetical order. We will refer to the videos of this database as the identifiers assigned to each



**Fig. 16** Average grayscale image for video Crowd<sub>0</sub>. **a** Original video, **b** stabilized video



**Fig. 17** Average image of the colored MEIs for video Crowd<sub>2</sub>. **a** Original video, **b** stabilized video

of them. Table 2 presents the second dataset, proposed by Liu et al. [23], which is composed of 139 videos divided into six categories. We will refer to the videos of this dataset as the name of the category followed by the identifier of each video assigned by the authors.

#### 4.1 Visual representation

Figure 7 presents the difference images by considering several indices within a sliding window for the originally unstable video. Figure 8 displays the images corresponding to the videos obtained after the stabilization process through the YouTube approach [15].

From Figs. 7 and 8, we can observe the occurrence of more white pixels in the difference images of the unstable video, indicating a greater amount of motion. This is even more visible with the increase in the frame distance. These results confirm that the use of the difference between frames with a certain distance can capture motion that is not perceived by comparison of adjacent frames.

Figure 9 shows the MEI for the same frame obtained for both the unstable and the stabilized videos. It can be verified that the MEI summarizes well the images of the differences and that the version of the stabilized video has darker pixels compared to that of the unstable video, which indicates the presence of less motion.

Figure 10 displays the gray-level average image of the MEIs for the unstable and stabilized video #4. From the figure, it is possible to observe an image with darker gray levels and with more defined shapes in the image corresponding to the stabilized video. Similar results were observed in all videos in the database under consideration. Figure 11 presents the histograms of the images shown in Fig. 10, where we can easily distinguish the image from the stabilized and non-stabilized video.

In the following results, we present the images obtained with the proposed method and compare them with the average grayscale of the video frames, as shown in Fig. 3 described in Sect. 2.

Figure 12 shows the color image of the average of the MEIs for the unstable and stabilized video #4. It is possible to observe a greater visual distinction when compared to the gray-level image. For the unstable video, the image contains red regions, which indicates the occurrence of a large amount of motion throughout the video. On the other hand, the image is predominantly blue and green for the stabilized video. Figure 13 shows the result obtained with the average grayscale for the same video. It is possible to notice that the stabilized version is better defined, whereas the unstable video image is more blurred. However, it is difficult to infer how much motion is present in the video from the image.

The drawback of the average grayscale image becomes even clearer in the comparison of the results obtained for the

video #7. Figures 14 and 15 show the results of the average grayscale and the average of the MEIs for video #7. From the gray-level image, it is not so easy to differentiate the unstable video from the stabilized one. In fact, the stabilized video seems to have more motion. On the other hand, the stabilized video presents an average MEI image with bluer tones, correctly indicating a smaller amount of motion.

The visual representation proposed in this work is efficient to show the amount of motion present in a video, making possible the evaluation and comparison of different stabilization methods. Our technique is more effective than the simple average of the gray levels of the video frames, which can generate inaccurate results when considering the intentional motion of the camera and small changes in the scene.

Figures 16 and 17 show the results obtained for the average grayscale image and the proposed visual representation in a video for a crowded scene.

From Figs. 16 and 17, we can see the differences between the image versions in the proposed visual representation, before and after the video stabilization. Even after stabilization, we can notice red color in the result, which is probably due to the presence of moving people in the scene. However, stronger tones of red are featured in the unstable version of the video, which characterizes a video in the presence of much motion. The images of the average grayscale, however, show little difference, demonstrating the superiority of our visual representation.

Figures 18 and 19 show the results obtained for the average grayscale image and the proposed visual representation in a video that contains parallax effect.

From Figs. 18 and 19, we can observe that redder tones were obtained in the unstable video version, whereas the image of the average grayscale presents little distinction between the two versions of the video.

Figures 20 and 21 illustrate the results obtained for the average grayscale image and the proposed visual representation in a video with fast translations.

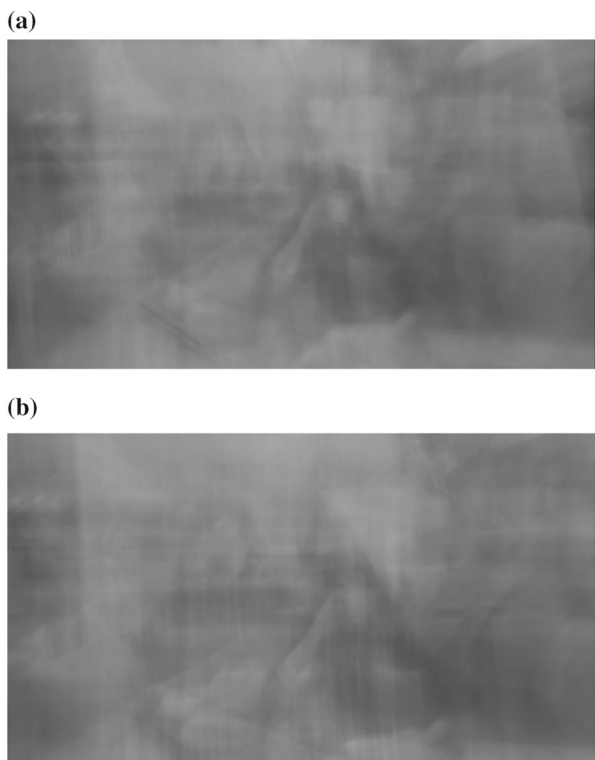
From Figs. 20 and 21, we can notice that a video in the presence of fast translations tends to have very red tones. Similarly to other cases, lighter tones are obtained in the stabilized version. After stabilization, the visual representation continues with red tones, since there is still a certain amount of motion desired in the video. Again, the visualization of the average grayscale image is not very effective.

Figures 22 and 23 present the results obtained for the average grayscale image and the proposed visual representation in a video with regular scene.

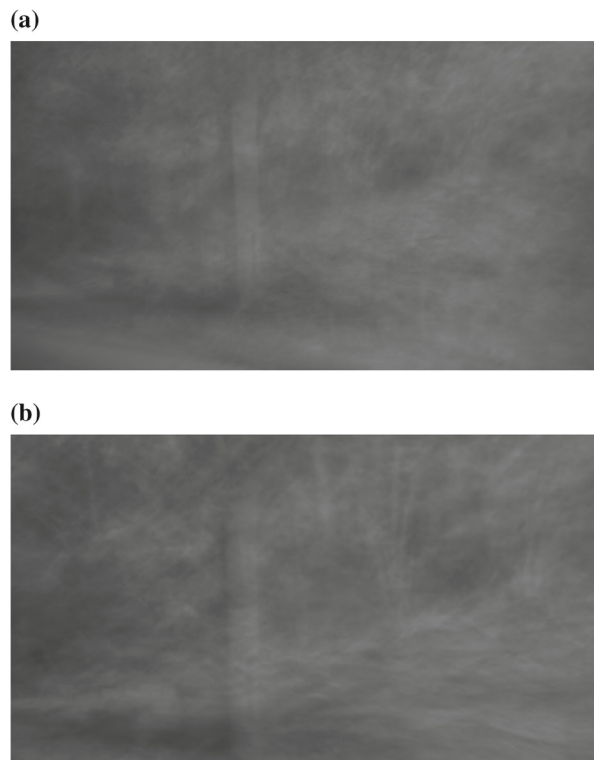
From Figs. 22 and 23, the image for the stabilized version has considerably lighter colors, once this scene has little movement. We can also notice that redder tones are present in the region where a person is moving.

Figures 24 and 25 show the results obtained for the average grayscale image and the proposed visual representation

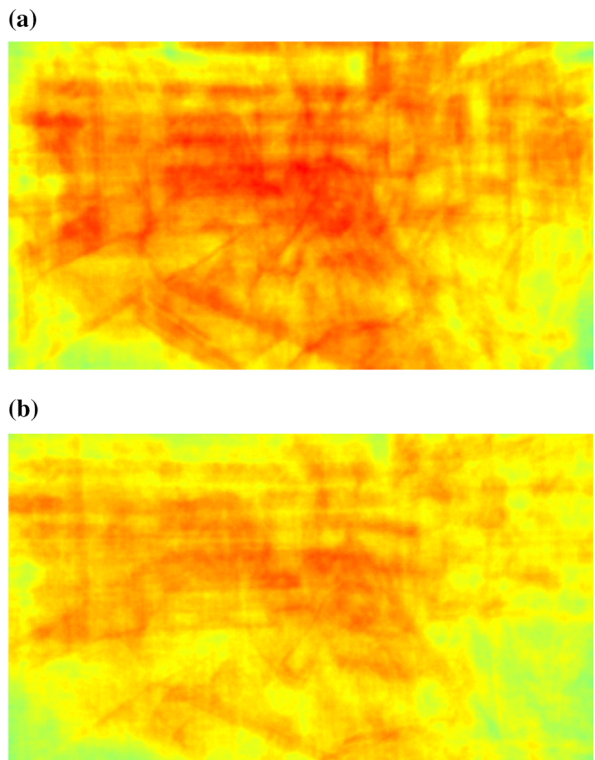




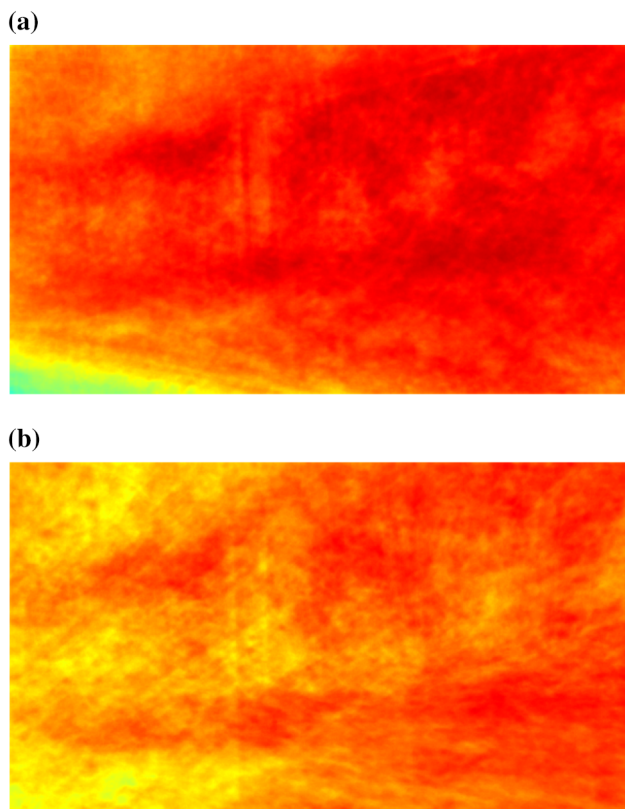
**Fig. 18** Average grayscale image for video Parallax<sub>0</sub>. **a** Original video, **b** stabilized video



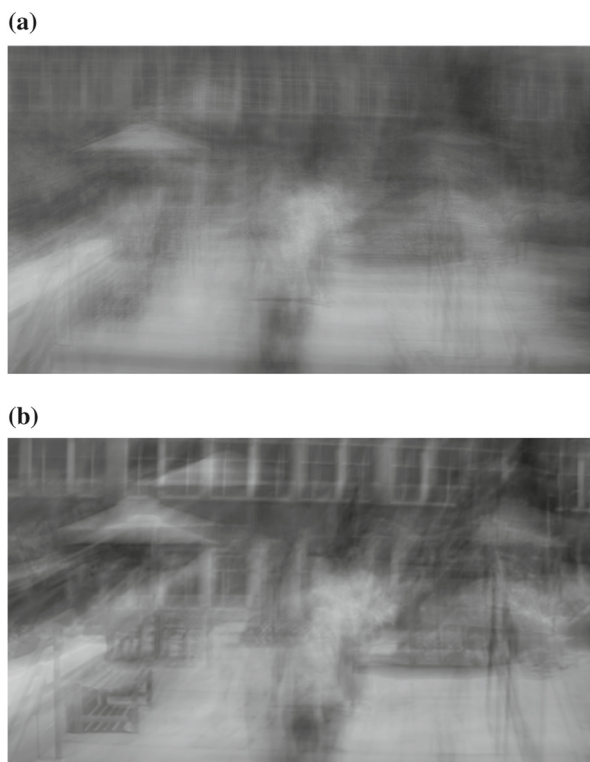
**Fig. 20** Average grayscale image of the average for video QuickRotation<sub>0</sub>. **a** Original video, **b** stabilized video



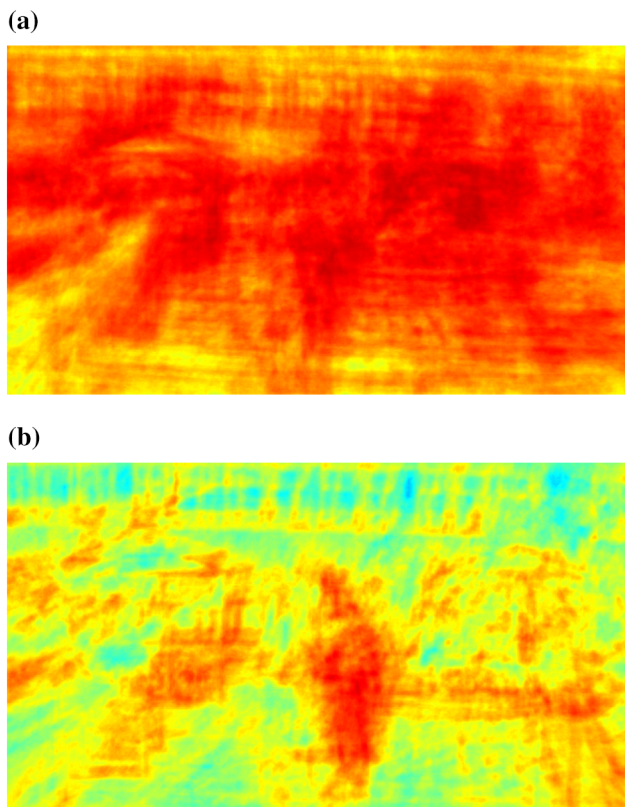
**Fig. 19** Average image of the colored MEIs for video Parallax<sub>0</sub>. **a** Original video, **b** stabilized video



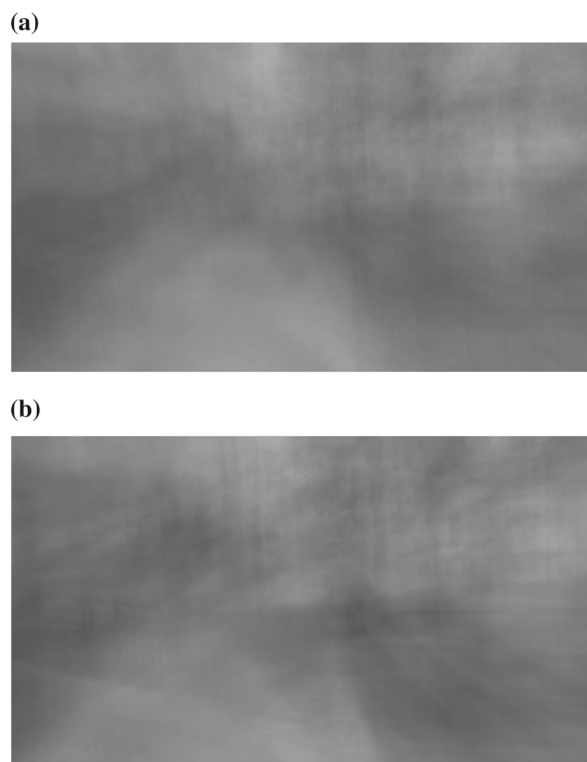
**Fig. 21** Average image of the colored MEIs for video QuickRotation<sub>0</sub>. **a** Original video, **b** stabilized video



**Fig. 22** Average grayscale image for video Regular<sub>0</sub>. **a** Original video, **b** stabilized video



**Fig. 23** Average image of the colored MEIs for video Regular<sub>0</sub>. **a** Original video, **b** stabilized video



**Fig. 24** Average grayscale image for video Running<sub>0</sub>. **a** Original video, **b** stabilized video

where the person shooting the video was running at the time of scene acquisition.

From Figs. 24 and 25, we can observe that the image tones are very reddish in both versions. This occurs due to the substantial change in the scene and to the motion caused by the person who shoots the video. Notwithstanding, we can notice lighter tones in the stabilized version.

Figures 26 and 27 present the results obtained for the average grayscale image and the proposed visual representation in a video in the presence of zoom.

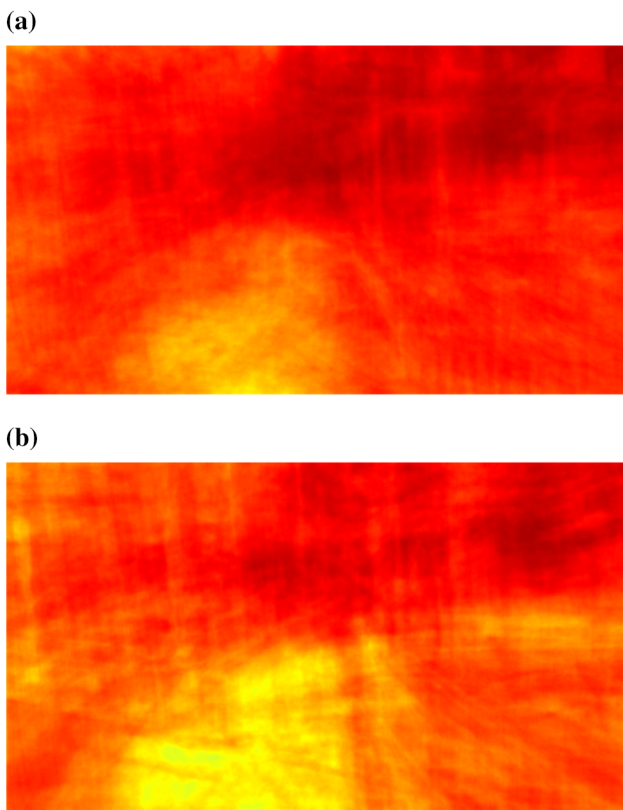
From Figs. 26 and 27, it is possible to observe that the stabilized version has lighter tones, which demonstrates the advantages of our method.

## 4.2 Objective metric

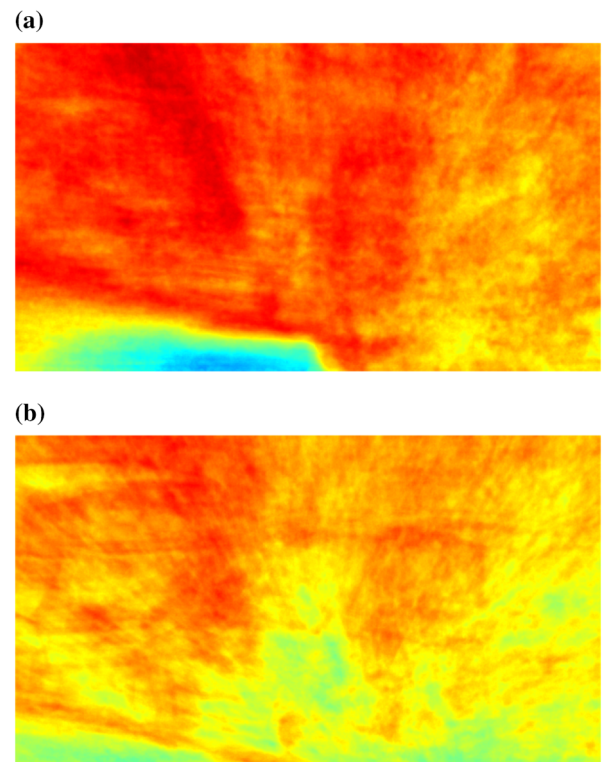
Table 3 displays the values of AAM, as well as the ITF values for the original videos and after the YouTube stabilization method [15].

Table 4 shows the AAM and ITF values for the tested videos before and after the stabilization process. Both versions are available in the database. The stabilized version was originally obtained with the method proposed by Liu et al. [23].

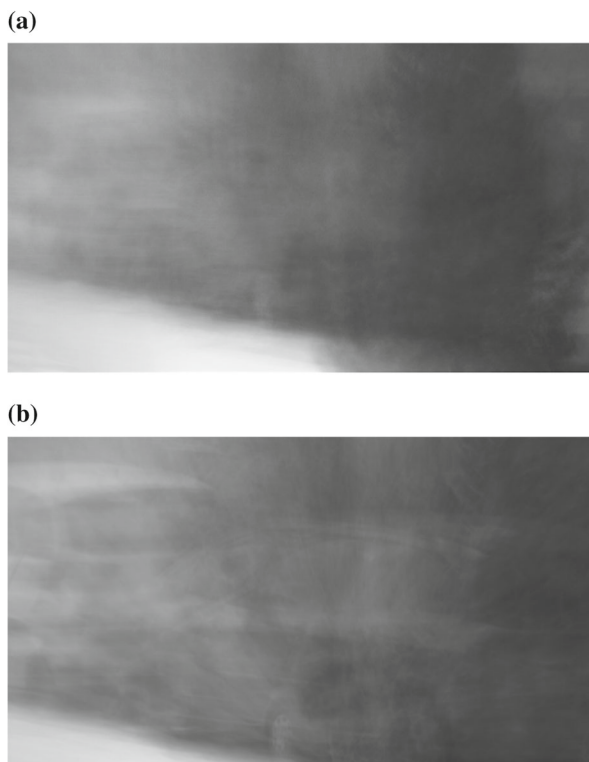
From Tables 3 and 4, it can be noticed that the proposed metric has consistent results with the ITF metric in the eval-



**Fig. 25** Average image of the colored MEIs for video *Running<sub>0</sub>*. **a** Original video, **b** stabilized video



**Fig. 27** Average image of the colored MEIs for video *Zooming<sub>0</sub>*. **a** Original video, **b** stabilized video



**Fig. 26** Average grayscale image for video *Zooming<sub>0</sub>*. **a** Original video, **b** stabilized video

**Table 3** AAM and ITF values for the videos from the first dataset

# Video	Original		YouTube	
	ITF	AAM	ITF	AAM
1	18.793	0.758	27.890	0.557
2	20.390	0.671	28.604	0.456
3	16.186	0.815	23.030	0.646
4	19.965	0.596	33.711	0.364
5	23.277	0.683	27.599	0.574
6	19.681	0.828	29.390	0.547
7	24.109	0.627	29.252	0.417
8	17.881	0.775	25.908	0.578
9	19.248	0.800	20.922	0.746
10	12.972	0.847	20.495	0.718
11	21.487	0.671	26.672	0.550
12	15.081	0.840	19.283	0.756
13	23.841	0.598	28.845	0.462
14	18.065	0.650	20.128	0.590
Mean	19.355	0.726	25.837	0.569

**Table 4** Mean AAM and ITF values for the videos from the second dataset

Category	Original		Stabilized	
	ITF	AAM	ITF	AAM
Crowd	19.479	0.787	23.699	0.731
Parallax	18.746	0.741	22.499	0.684
Quick rotation	19.963	0.758	24.166	0.705
Regular	19.468	0.713	28.248	0.567
Running	17.326	0.828	22.765	0.750
Zooming	20.113	0.746	24.630	0.684
Mean	19.183	0.762	24.334	0.687

**Table 5** AAM and ITF values for video #4

Gaussian $\sigma = 10$		Gaussian $\sigma = 40$		Gaussian $\sigma = 890$	
ITF	AAM	ITF	AAM	ITF	AAM
26.840	0.488	33.380	0.385	33.865	0.375

uated videos, which demonstrates that it can be used as an alternative to the ITF. From Table 4, we can see that the mean value of AAM is smaller in Regular, Zooming and Parallax categories, which have a lower amount of movement in their videos.

Table 5 presents the values of AAM and ITF metrics for video #4 stabilized through a simple method, where a Gaussian smoothing filter is applied with different values of  $\sigma$ .

From Table 5, it can be seen that the ITF and AAM values decrease with the increase of  $\sigma$ . This occurs because the method considered the motion as undesired and corrected most of the motion with increasing  $\sigma$ . It is possible to observe that, with  $\sigma = 890$ , the ITF obtained with the Gaussian filter is superior to that obtained with YouTube method. However, the video generated with the Gaussian filter is visually more unstable, containing several distortions.

The AAM values, also reported in Table 4, are not smaller than the value obtained with the YouTube method and, therefore, more consistent with the visual result of the video.

## 5 Conclusions and future work

This work presented a novel visual representation technique based on the motion energy image (MEI) for the subjective evaluation of video stabilization. The representation was constructed from the mean of MEIs calculated for all the video frames, and then highlighted with a pseudocolor transformation. In addition, the average gray-level of the representation, denoted average amount of motion (AAM), was proposed as a new objective metric.

We were able to characterize the amount of spatial motion, as well as its location, present in the video. Assuming an unstable video has greater amount of motion than its stabilized version, we can employ this technique to evaluate the video stabilization process.

The results showed that the proposed visual representation is adequate and expresses well both the amount and location of spatial motion. We compared our representation to the mean gray-level frames in several different scenarios and verified that the representation performed better.

The proposed objective metric presented consistent results. In some cases, the AAM overcame the Interframe Transformation Fidelity (ITF), which is the most commonly used objective metric to evaluate video stabilization methods.

As directions for future work, we intend to conduct experiments with participation of people to validate the proposed visual representation. We also plan to investigate the direction and speed of intensity changes in the video frames through visual rhythms for the subjective evaluation of video stabilization. Finally, we intend to investigate other objective metrics obtained from the visual representation technique proposed in this work, as well as its use in conjunction with visual rhythms for the characterization and evaluation of video stabilization.

**Acknowledgements** The authors are thankful to São Paulo Research Foundation (FAPESP Grants #2017/12646-3 and #2014/12236-1) and National Council for Scientific and Technological Development (CNPq Grant #305169/2015-7) for their financial support.

## References

- Ahad, M.A.R.: Motion History Images for Action Recognition and Understanding. Springer, Berlin (2012)
- Amanatiadis, A.A., Andreadis, I.: Digital image stabilization by independent component analysis. *IEEE Trans. Instrum. Meas.* **59**(7), 1755–1763 (2010)
- Battiato, S., Gallo, G., Puglisi, G., Scellato, S.: SIFT features tracking for video stabilization. In: 14th International Conference on Image Analysis and Processing, pp. 825–830. IEEE (2007)
- Borgo, R., Chen, M., Daubney, B., Grundy, E., Heidemann, G., Höferlin, B., Höferlin, M., Leitte, H., Weiskopf, D., Xie, X.: State of the art report on video-based graphics and video visualization. *Comput. Graph. Forum* **31**, 2450–2477 (2012)
- Chang, H.C., Lai, S.H., Lu, K.R.: A robust and efficient video stabilization algorithm. *IEEE Int. Conf. Multimed. Expo IEEE* **1**, 29–32 (2004)
- Chang, J.Y., Hu, W.F., Cheng, M.H., Chang, B.S.: Digital image translational and rotational motion stabilization using optical flow technique. *IEEE Trans. Consum. Electron.* **48**(1), 108–115 (2002)
- Chen, B., Zhao, J., Wang, Y.: Research on evaluation method of video stabilization. In: International Conference on Advanced Material Science and Environmental Engineering, pp. 253–258 (2016)
- Chen, B.H., Kopylov, A., Huang, S.C., Seredin, O., Karpov, R., Kuo, S.Y., Lai, K.R., Tan, T.H., Gochoo, M., Bayanduuren, D.: Improved global motion estimation via motion vector clustering for video stabilization. *Eng. Appl. Artif. Intell.* **54**, 39–48 (2016)

9. Chen, B.Y., Lee, K.Y., Huang, W.T., Lin, J.S.: Capturing intention-based full-frame video stabilization. *Comput. Graph. Forum* **27**, 1805–1814 (2008)
10. Choi, S., Kim, T., Yu, W.: Robust video stabilization to outlier motion using adaptive RANSAC. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1897–1902. IEEE (2009)
11. Cirne, M.V.M., Pedrini, H.: A video summarization method based on spectral clustering. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 479–486. Springer, Berlin (2013)
12. Cirne, M.V.M., Pedrini, H.: Summarization of videos by image quality assessment. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 901–908. Springer, Berlin (2014)
13. Ertürk, S.: Real-time digital image stabilization using Kalman filters. *Real Time Imaging* **8**(4), 317–328 (2002)
14. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, Upper Saddle River (2002)
15. Grundmann, M., Kwatra, V., Essa, I.: Auto-directed video stabilization with robust L1 optimal camera paths. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 225–232 (2011)
16. Huang, T.S.: *Image Sequence Analysis*, vol. 5. Springer, Berlin (2013)
17. Jia, R., Zhang, H., Wang, L., Li, J.: Digital image stabilization based on phase correlation. *Int. Conf. Artif. Intell. Comput. Intell.* **IEEE 3**, 485–489 (2009)
18. Joshi, N., Kienzle, W., Toelle, M., Uyttendaele, M., Cohen, M.F.: Real-time hyperlapse creation via optimal frame selection. *ACM Trans. Graph.* **34**(4), 63 (2015)
19. Ko, S.J., Lee, S.H., Lee, K.H.: Digital image stabilizing algorithms based on bit-plane matching. *IEEE Trans. Consum. Electron.* **44**(3), 617–622 (1998)
20. Kumar, S., Azartash, H., Biswas, M., Nguyen, T.: Real-time affine global motion estimation using phase correlation and its application for digital image stabilization. *IEEE Trans. Image Process.* **20**(12), 3406–3418 (2011)
21. Lin, C.T., Hong, C.T., Yang, C.T.: Real-time digital image stabilization system using modified proportional integrated controller. *IEEE Trans. Circuits Syst. Video Technol.* **19**(3), 427–431 (2009)
22. Litvin, A., Konrad, J., Karl, W.C.: Probabilistic video stabilization using Kalman filtering and mosaicing. In: *Electronic Imaging, International Society for Optics and Photonics*, pp. 663–674 (2003)
23. Liu, S., Yuan, L., Tan, P., Sun, J.: Bundled camera paths for video stabilization. *ACM Trans. Graph.* **32**(4), 78 (2013)
24. Lowe, D.G.: Object recognition from local scale-invariant features. *Seventh IEEE Int. Conf. Comput. Vis.* **IEEE 2**, 1150–1157 (1999)
25. Marcenaro, L., Vernazza, G., Regazzoni, C.S.: Image stabilization algorithms for video-surveillance applications. *Int. Conf. Image Process.* **IEEE 1**, 349–352 (2001)
26. Matsushita, Y., Ofek, E., Ge, W., Tang, X., Shum, H.Y.: Full-frame video stabilization with motion inpainting. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(7), 1150–1163 (2006)
27. Morimoto, C., Chellappa, R.: Fast electronic digital image stabilization. In: *13th International Conference on Pattern Recognition*, vol. 3, pp. 284–288. IEEE (1996)
28. Niskanen, M., Silvén, O., Tico, M.: Video stabilization performance assessment. In: *IEEE International Conference on Multimedia and Expo*, pp. 405–408. IEEE (2006)
29. Puglisi, G., Battiato, S.: A robust image alignment algorithm for video stabilization purposes. *IEEE Trans. Circuits Syst. Video Technol.* **21**(10), 1390–1400 (2011)
30. Qu, H., Song, L., Xue, G.: Shaking video synthesis for video stabilization performance assessment. In: *Visual Communications and Image Processing*, pp. 1–6. IEEE (2013)
31. Ratakonda, K.: Real-time digital video stabilization for multimedia applications. *IEEE Int. Symp. Circuits Syst.* **IEEE 4**, 69–72 (1998)
32. Ryu, Y.G., Chung, M.J.: Robust online digital image stabilization based on point-feature trajectory without accumulative global motion estimation. *IEEE Signal Process. Lett.* **19**(4), 223–226 (2012)
33. Schoeffmann, K., Lux, M., Taschwer, M., Boeszöermyeni, L.: Visualization of video motion in context of video browsing. In: *IEEE International Conference on Multimedia and Expo*, pp. 658–661. IEEE (2009)
34. Shen, Y., Guturu, P., Damarla, T., Buckles, B.P., Namuduri, K.R.: Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework. *IEEE Trans. Consum. Electron.* **55**(3), 1714–1721 (2009)
35. Shukla, D., Jha, R.K.: A robust video stabilization technique using integral frame projection warping. *Signal Image Video Process.* **9**(6), 1287–1297 (2015)
36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
37. Yang, J., Schonfeld, D., Mohamed, M.: Robust video stabilization based on particle filter tracking of projected camera motion. *IEEE Trans. Circuits Syst. Video Technol.* **19**(7), 945–954 (2009)
38. Zheng, Q., Yang, M.: A video stabilization method based on inter-frame image matching score. *Glob. J. Comput. Sci. Technol.* **17**(1), 1–6 (2017)



**Marcos Roberto e Souza** received his B.Sc. degree in Computer Science from Federal University of Technology—Parana, Brazil. He received his M.Sc. degree in Computer Science from University of Campinas, Brazil. He is currently a Ph.D. student in the Institute of Computing at the University of Campinas, Brazil. His research interests include image analysis, machine learning, computer vision, and video processing.



**Helio Pedrini** is currently a professor in the Institute of Computing at the University of Campinas, Brazil. He received his Ph.D. degree in Electrical and Computer Engineering from Rensselaer Polytechnic Institute, Troy, NY, USA. He received his M.Sc. in Electrical Engineering and his B.Sc. in Computer Science, both degrees from the University of Campinas, Brazil. His research interests include image processing, computer vision, pattern recognition, machine learning, computer graphics, and scientific visualization.