RESEARCH **Open Access**

Visual rhythms for qualitative evaluation of video stabilization

Marcos Roberto e Souza[†] and Helio Pedrini[*†] (iD)

*Correspondence:
helio@ic.unicamp.br
[†]Marcos Roberto e Souza and Helio Pedrini contributed equally to this work.
Institute of Computing, University of Campinas, Av. Albert Einstein 1251, Campinas, 13083-852 Brazil

Abstract

Recent technological advances have enabled the development of compact and portable cameras for the generation of large volumes of video content. Several applications have benefited from such significant growth of multimedia data, such as telemedicine, surveillance and security, entertainment, teaching, and robotics. However, videos captured by amateurs are subject to unwanted motion or vibration while handling the camera. Video stabilization techniques aim to detect and remove glitches or instabilities caused during the acquisition process to enhance visual quality. In this work, we introduce and analyze a novel representation based on visual rhythms for qualitative evaluation of video stabilization methods. Experiments conducted on different video sequences are performed to demonstrate the effectiveness of the visual representation as qualitative measure for evaluating video stability. In addition, we present a proposal to calculate an objective metric extracted from the visual rhythms.

Keywords: Video stabilization, Qualitative evaluation, Visualization, Video processing

1 Introduction

The popularization of mobile devices in recent years has contributed to making video acquisition possible for a variety of applications. Handling such devices generally causes unwanted motion during the video generation, which inevitably affects the quality of the final video.

Video stabilization [1–15] aims to remove undesired motion in camera handling during video acquisition. Efficient methods for stabilization of videos are important to improve their quality according to human perception or to facilitate certain tasks, such as multimedia indexing and retrieval [16–18].

Techniques and metrics for quality evaluation must be well established so that video stabilization approaches can be developed, refined, and compared in a consistent manner. Therefore, ineffective evaluation measures may lead to the development of inadequate techniques, compromising the advance of state-of-the-art video stabilization approaches.

Most of the quantitative techniques for the evaluation of video stabilization available in the literature are inaccurate and, in some cases, incompatible with human visual perception. Moreover, the techniques used to evaluate and report the results subjectively are

little explored. In this work, we introduce and evaluate the use of visual rhythms as a novel mechanism for the qualitative evaluation of video stabilization methods.

Experimental results demonstrate that the visual rhythms are effective to evaluate the stability of camera motion by differentiating stable and unstable videos. Furthermore, it allows to determine how and when a given motion occurs. More complex types of motion, such as zoom and quick shifts, can also be identified.

This paper is organized as follows. Relevant concepts and related work are briefly described in Section 2. The use of visual rhythms for subjective evaluation of video stabilization is presented in Section 3. Experimental results are reported and discussed in Section 4. Final remarks and directions for future work are outlined in Section 5.

2 Background

Different categories of stabilization systems have been proposed to improve the quality of videos. The three most common types are mechanical stabilization, optical stabilization, and digital stabilization.

Mechanical video stabilization typically uses sensors to detect camera shifts and compensate for undesired movements. A common way is to use gyroscopes to detect motion and send signals to motors connected to small wheels, such that the camera can move in the opposite direction of motion.

Optical video stabilization [19] is widely used in photographic cameras and consists of a mechanism to compensate for the angular and translational movements of the cameras, stabilizing the image before it is recorded on the sensor. A form of optical stabilization introduces a gyroscope to measure velocity differences at distinct instants to distinguish between normal and undesired motion.

Digital video stabilization is implemented in software without the use of special devices. Digital video stabilization methods are commonly categorized into two-dimensional (2D) and three-dimensional (3D) approaches. In the first category, techniques estimate camera motion from two consecutive frames and apply 2D transformations to stabilize the video. In the second category, techniques reconstruct the camera trajectories from 3D transformations [20, 21], such as scaling, translation, and rotation.

In the context of image and video processing, the evaluation can be classified as (i) objective, when obtained through functions applied between two images [22] or video frames, and (ii) subjective, when the analysis is performed by human observers. In both cases, a desired goal is to assess stabilization based on criteria in agreement with the perception of the human visual system.

2.1 Objective evaluation

Criteria for measuring the amount and nature of the camera displacement have been proposed to evaluate the quality of video stabilization in an objective manner [23]. 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. The divergence is also verified, which indicates the amount of expected displacement. For an overall assessment, the blurring caused by the stabilization process is considered.

Most of the video stabilization approaches found in the literature have adopted the interframe transformation fidelity (ITF) [24–28], which can be expressed as the peak

signal-to-noise ratio (PSNR) of video frames. More recent approaches have considered the structural similarity (SSIM) [29] as an alternative to PSNR [28].

Liu et al. [30] employed the amount of energy present in the low-frequency portion of the 2D motion estimated as a stability metric. The rate of frame cropping and distortion are used to assess the stabilization process more generally.
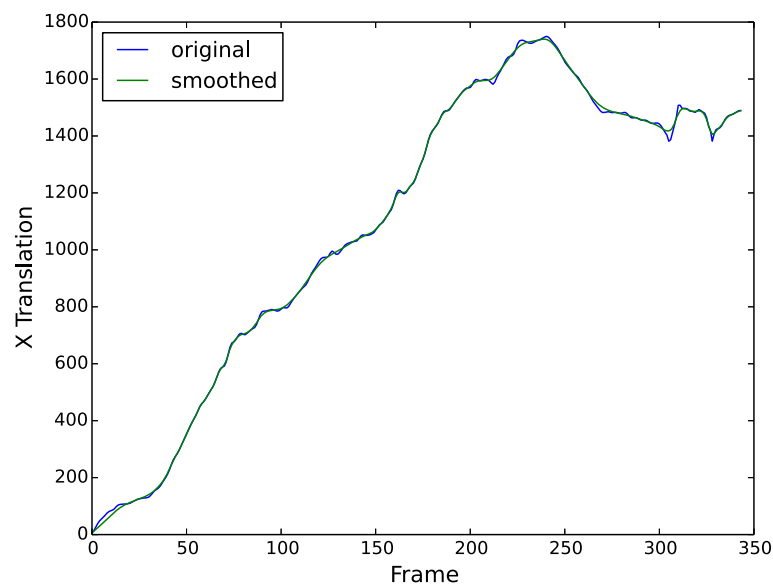
Synthesizing unstable videos from stable videos has been proposed for the evaluation of video stabilization [31] in order to provide the ground-truth of the stable videos. The methods are evaluated according to two aspects: (i) the distance between the stabilized frame and the reference frame and (ii) the average of the SSIM between each pair of consecutive frames.

Due to the weaknesses of ITF in motion videos, an evaluation method based on the variation of the intersection of angles between the global motion vectors, calculated from the scale-invariant feature transform (SIFT) keypoints [32], was proposed to evaluate the video stabilization process [33]. In fixed-camera videos, the ITF is considered, however, only for overlapping the frame background, instead of the entire frame.
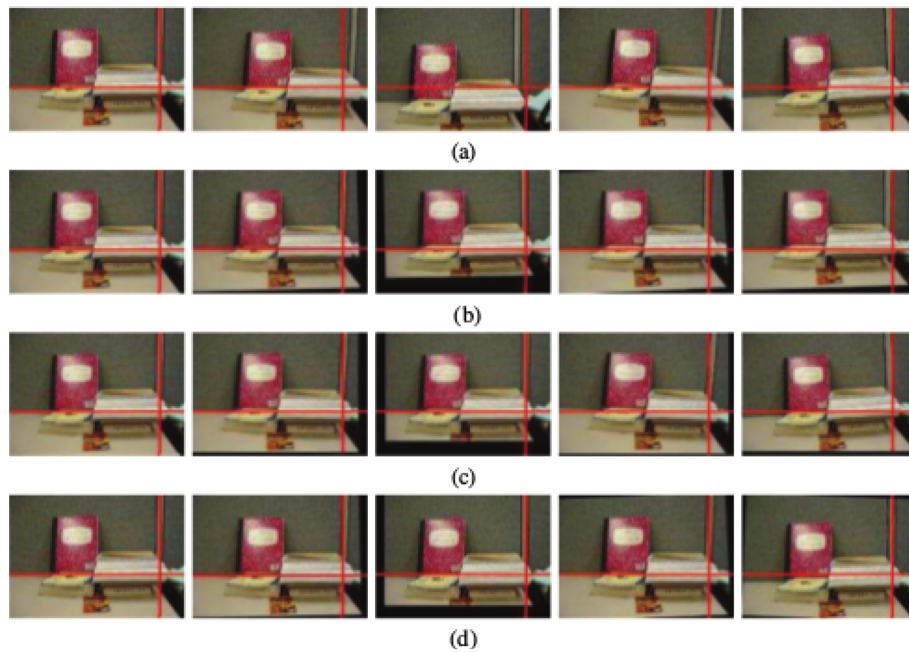
2.2 Subjective evaluation

Several methods found in the literature briefly describe and analyze review the trajectories made by the camera and the trajectories of the stabilized video [34–38]. These trajectories are usually related to the different factors that compose the estimated 2D motion. For instance, the approaches present the camera path for horizontal and vertical translations and rotations. Figure 1 shows an example of path for horizontal translation estimated from the original (blue) and smoothed (green) trajectory.

From the trajectory, it is possible to identify when a motion occurs and its intensity in the original video, as well as such motion after its smoothing. This type of visualization can be very useful to analyze the behavior of the motion smoothing step used in a certain method. However, its result depends on the technique used in estimating the motion, so

Fig. 1 Horizontal translation path of a camera

Fig. 2 Sequence of video frames. **a** Original video. **b**–**d** Different versions of the stable video. Extracted from [40]

that the trajectory does not reliably represent the video motion. Thus, the trajectory may not be a good alternative to the evaluation of the stabilization quality, as well as not an adequate visualization for videos with spatially distinct motion.
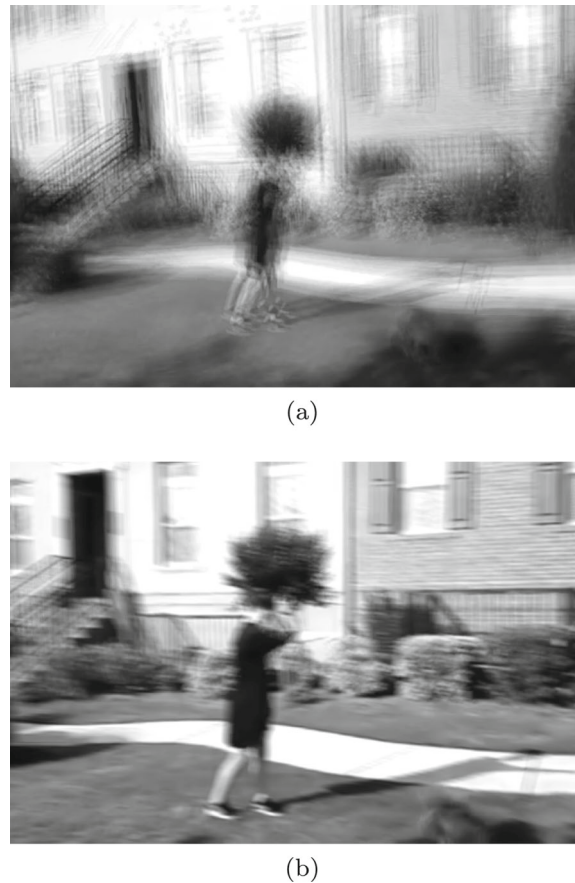
Some approaches in the literature deal with frame sequences usually superimposed by horizontal and vertical lines [25, 28, 35–37, 39, 40]. Thus, it is possible to check the alignment of a small set of consecutive frames. Figure 2 illustrates an example of such type of visualization, where objects intercepted by lines are more aligned in the stabilized video.

From the sequence of frames, the displacement of each frame is noticeable, 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 is considered, compromising the analysis of the entire video.

Furthermore, there are approaches that summarize a video in a single image calculated through the average gray levels of the frames [41, 42], as shown in Fig. 3. Better-defined images are expected for more stable videos. From this representation, it is possible to check if the video has more amount of motion; however, it is difficult to determine the nature of video motion.

In a broader context, video visualization is concerned with the creation of a new visual representation, obtained from an input video, capable of indicating its characteristics and important events [43]. Video visualization techniques can generate different types of output data, such as another video, a collection of images or a single image. Borgo et al. [43] 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, motion histograms were proposed in the HSV color space [44]. Motion histograms are obtained by means of motion vectors contained in H.264/AVC codecs.
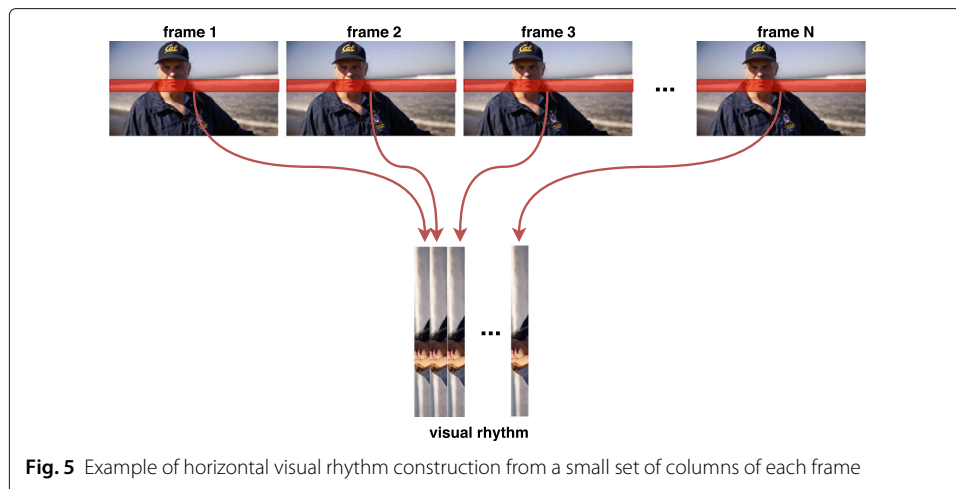
Fig. 3 The average gray levels for the first ten frames. **a** Original video. **b** Stabilized video

Figure 4 presents an example of the visualization, where each frame of the video is represented by a vertical line, such that the motion direction is mapped by different colors and the motion intensity by brightness values. As a disadvantage, this technique suffers from the presence of noise in the motion vectors, introduced by the motion estimation algorithm [44].

Visual rhythm [45] (VR) corresponds to a summary of temporal information of a video represented as a single image. This is done by concatenating portions of information from each frame of the video. Visual rhythms have been generally applied in the context of video identification and classification, for instance, location of video subtitles, recognition of person action detection of video shot boundaries, detection of face spoofing, among others [46–50]. Unlike these approaches, the visual rhythms are used in this work to create a representation of temporal information that allows the evaluation of the video stabilization by humans.

Fig. 4 Histograms of motion with HSV color space. Extracted from [44]

Fig. 5 Example of horizontal visual rhythm construction from a small set of columns of each frame

Typically, two different paths for constructing the visual rhythms are considered when traversing each video: horizontal and vertical. Such representations differ according to the information that is extracted from the video frames. The vertical rhythm extracts the information from the columns of each frame, whereas the horizontal rhythm is constructed from the rows of each frame.

A single column or row (or a small set of them) of each frame is usually used to construct the visual rhythm. Figure 5 illustrates the construction of a horizontal visual rhythm, as commonly described in the literature. However, the construction of a visual rhythm is very susceptible to different strategies for video traversal, for instance, a zigzag path, where an alternating direction might extract patterns from the video frames more appropriately for a certain problem.

3   Methods

In this work, the visual rhythms are constructed by traversing the video at vertical and horizontal directions. However, as opposed to using a single row or column (or a small set of rows or columns), we use the average of the columns for the vertical rhythm and the average of the rows for the horizontal rhythm.

For both path directions, the rhythm is obtained from the sequential concatenation of the intensity values, such that the jth column of the visual rhythm image corresponds to the intensity values in the jth frame. In the horizontal rhythm, a rotation is performed on the rows in order to obtain the columns in the final image. The width of a visual rhythm corresponds to the number of video frames, whereas its height corresponds to the height or width of the frames for the vertical or horizontal rhythm, respectively.

Figure 6 shows the relations between the pixels of the neighborhood in a visual rhythm image, from which we can see that the visual rhythm maintains the temporal and spatial information of the video. Thus, the temporal behavior of the gray levels in a certain region can be easily visualized. This provides information on how and when movements occur in the video, that is, in addition to being able to distinguish the direction, the intensity, and the form that the movements are spatially arranged, we can verify the frequency of certain type of movement and determine the moments of its occurrence. Stable video is expected to have a more uniform visual rhythm, with fewer twitches and better defined
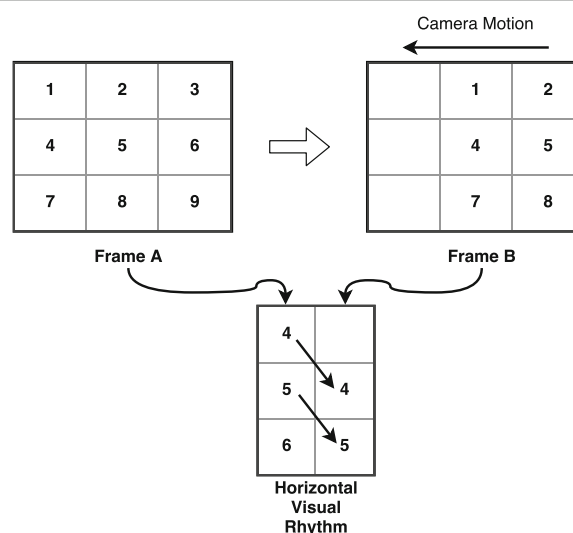
Neighbor $i-1$ of the previous frame	Neighbor $i-1$	Neighbor $i-1$ of the next frame
Same pixel i of the previous frame	Pixel i	Same pixel i of the next frame
Neighbor $i+1$ of the previous frame	Neighbor $i+1$	Neighbor $i+1$ of the next frame

Fig. 6 Patterns for pixel neighborhood in the visual rhythm

curves. We refer to "neighbor $i-1$" as the pixel that is on the row immediately above the row of pixel i in the column that represents information extracted from a frame, whereas "neighbor $i+1$" corresponds to the pixel immediately below the row of pixel i.

Figure 7 shows the construction of a horizontal rhythm for two 3×3 frames. At the transition between frames A and B, the camera moves from right to left, causing the pixels to be to the right of their original position. Thus, when obtaining the horizontal rhythm, the pixels of the column corresponding to frame B are below the equivalent pixels of frame A, thereby forming a declination.

The separation of the vertical and horizontal visual rhythms is important to thoroughly detect and evaluate problems in the video stabilization process. From the vertical rhythm, we can analyze the characteristics of the motion in the y axis. Thus, inclined rhythm lines indicate camera movements from the bottom to top, whereas declined lines indicate camera movements from top to bottom. From the horizontal rhythm, in turn, we have the characteristics of the motion in the x axis. Thus, sloped lines indicate camera movements from left to right, whereas declined lines indicate camera movement from right to left.

Fig. 7 Direction of horizontal visual rhythm

The use of only one column or row in the extraction of information from each frame may be inadequate since it considers little information of the frame. In addition, it makes horizontal and vertical separation less accurate. This problem can be seen in Fig. 8, where a vertical movement of the camera occurs, which can influence the horizontal rhythm, depending on the difference of the pixels between the rows. Thus, the average of the columns or rows is adopted in our work to compensate for this difference, making the horizontal rhythm less sensitive to vertical movements, and the vertical rhythm less sensitive to horizontal movements.
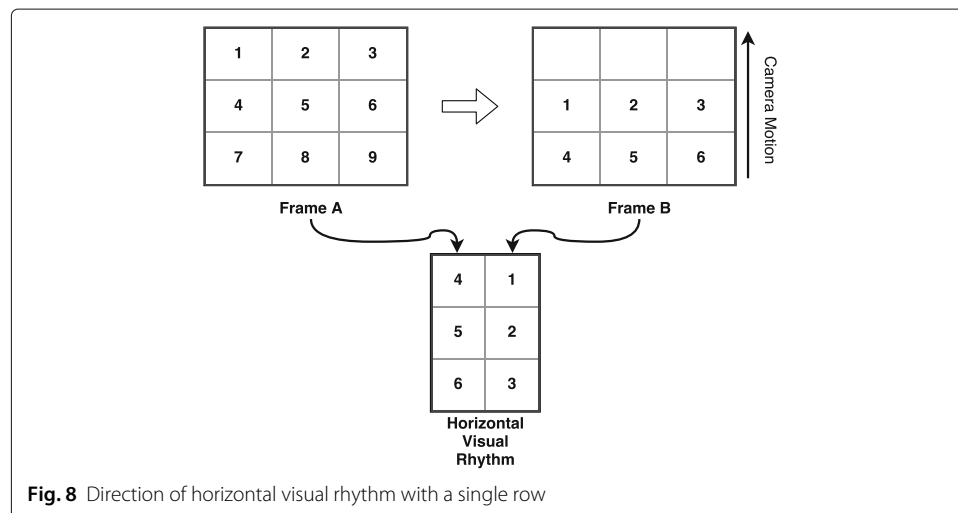
In Fig. 8, both columns of the horizontal rhythm should have either the same or very close values. However, with a single row in each frame, the direction of the rhythm is uncertain.

As post-processing, we apply an adaptive histogram equalization technique through the contrast-limited adaptive histogram equalization (CLAHE) [51]. This is done to improve the contrast of the visual rhythm, facilitating human perception.

The construction of the visual rhythms is not based on motion estimation, as occurs in other visualizations, shown in Section 2. Therefore, their performance is not dependent on any motion estimation technique, which makes the representation of the video motion more reliable. In the context of video stabilization, such independence of methods for motion estimation is crucial to allow a more unbiased assessment of the results.

The complexity of constructing a visual rhythm depends on three main factors: width W of the video frames, height H of the frames, and number N of frames in the video. To calculate an average row in the construction of a horizontal visual rhythm, we need to compute W averages. The calculation of each mean considers H values. Thus, we have $\theta(WH)$ as the asymptotic complexity of constructing an average row. The same complexity is taken for the computation of an average column in a vertical visual rhythm. Since either a row or a column should be calculated for each frame of the video, we have $\theta(WHN)$ as the final complexity for constructing a visual rhythm.

Among the good practices in the construction of visual rhythms for the evaluation of video stabilization results, we recommend the following:

Fig. 8 Direction of horizontal visual rhythm with a single row

- Crop the frames of the stabilized video so that there are no pixels with null information (since null information may imply inadequate row or column averages);
- Preserve the frame rate of the video in order to not change its number of frames or generate visual rhythms of different sizes;
- Rescale the video frames to the original size in order for the visual rhythms to have the same size.

3.1 Insights into objective metrics

This subsection provides some insights into the calculation of objective metrics from the visual rhythms for the evaluation of the video stabilization process. It is important to mention that we do not intend to replace existing objective metrics in the literature with the proposed objective metric, but to show that a metric can be extracted to distinguish unstable from stable videos.
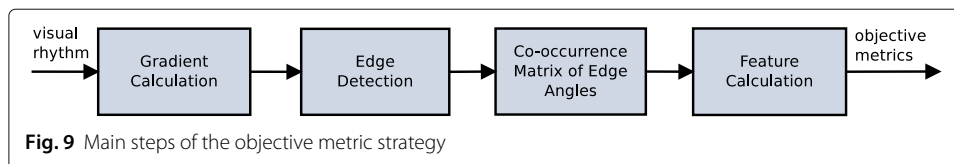
In the visual rhythm, the behavior of the movement present in the video is represented by the shapes of the curves. A more stable video has rhythms with smoother curves. As shown in Fig. 7, the directions of the visual rhythm can be observed in each column pair of pixels. Objective metrics can be calculated from the texture of visual rhythms. We conjecture that a softer visual rhythm has more regular directions, with less abrupt changes in the near directions. Thus, to obtain a new objective metric from the visual rhythm, the directions and their changes must be computed. Figure 9 illustrates the strategy for calculating the metric.

Initially, we calculate the visual rhythm gradients in order to obtain the directions of each pixel of the rhythm. This was implemented through the Sobel filter [52]. The gradients are decomposed into magnitude and angle information.

A thresholding with the Otsu algorithm [53] is applied to the magnitude values to determine the edges of the visual rhythm. This is done in order to consider only the edge angles in the following calculations. Then, a co-occurrence representation is calculated based on the gray level co-occurrence matrix (GLCM) [54]. However, it considers the co-occurrence of the angles of the edges in the direction of the angles themselves.

Initially, we eliminated the sign from the angles, leaving them in the range of 0 to 180°. For the calculation of the co-occurrence matrix M, we consider n directions $D = \{d_0, d_1, ..., d_n\}$, resulting in a matrix of size $n \times n$. The angles are then quantized in possible directions. For each pixel i belonging to the edge, we have its angle $\theta_i \in D$, from which we calculate the closest pixel j in the direction of θ_i. Then, it counts as a co-occurrence at position M_{θ_i, θ_j}, that is, an increment at M_{θ_i, θ_j}. For cases where θ_i are different from the important angles, we have two pixels j_1 and j_2. Thus, the two positions of the matrix are incremented proportionally to the distances of the angles.

Finally, the matrix is normalized by the sum of its elements. Thus, the value of the matrix at position M_{θ_i, θ_j} indicates the probability that θ_j is the next direction of the visual rhythm, since the previous one was θ_i. From the co-occurrence matrix generated, we can

Fig. 9 Main steps of the objective metric strategy

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	lf_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

calculate features to obtain objective metrics. Among the textural features defined by Haralick and Shanmugam [54], the homogeneity can be expressed as

$$\text{homogeneity} = \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{1}{1 + (i-j)^2} M_{i,j} \qquad (1)$$

The homogeneity feature, when calculated from the co-occurrence matrix of the edge angles, will assume larger values the closer the angles of consecutive directions.

Several other measures could be developed to extract useful information to qualify the stabilization from their visual rhythms. For this, a thorough investigation is necessary to identify which aspects are important to characterize an unstable motion and how to obtain such aspects through visual rhythm. These tasks may involve both handcrafted features and machine learning techniques.

4   Results and discussion

This section describes and evaluates the experimental results obtained with two datasets. All the videos considered in our experiments were obtained from two publicly available databases: GaTech VideoStab[1] [55] and the database proposed by Liu et al.[2] [30].

Table 1 reports a summary of the first database with videos in alphabetical order. We will refer to the videos in this database through the identifiers assigned to each of them. Table 2 presents the database proposed by Liu et al. [30], which is divided into six categories, containing a total of 139 videos. We will refer to the videos in this dataset by the name of the category followed by the identifier of each video, attributed by the authors. Due to space limitations, we report only a few visual rhythms that illustrate the results obtained from these databases, which have been confirmed in the other videos.

Figure 10 presents the visual rhythms generated for the video #12 before and after the video stabilization process. In order to obtain the stabilized version of the video, we submit it to `YouTube`, which applies one of the state-of-the-art digital video stabilization approaches [55]. The width of all the images presented in this section was considered constant for a better organization.
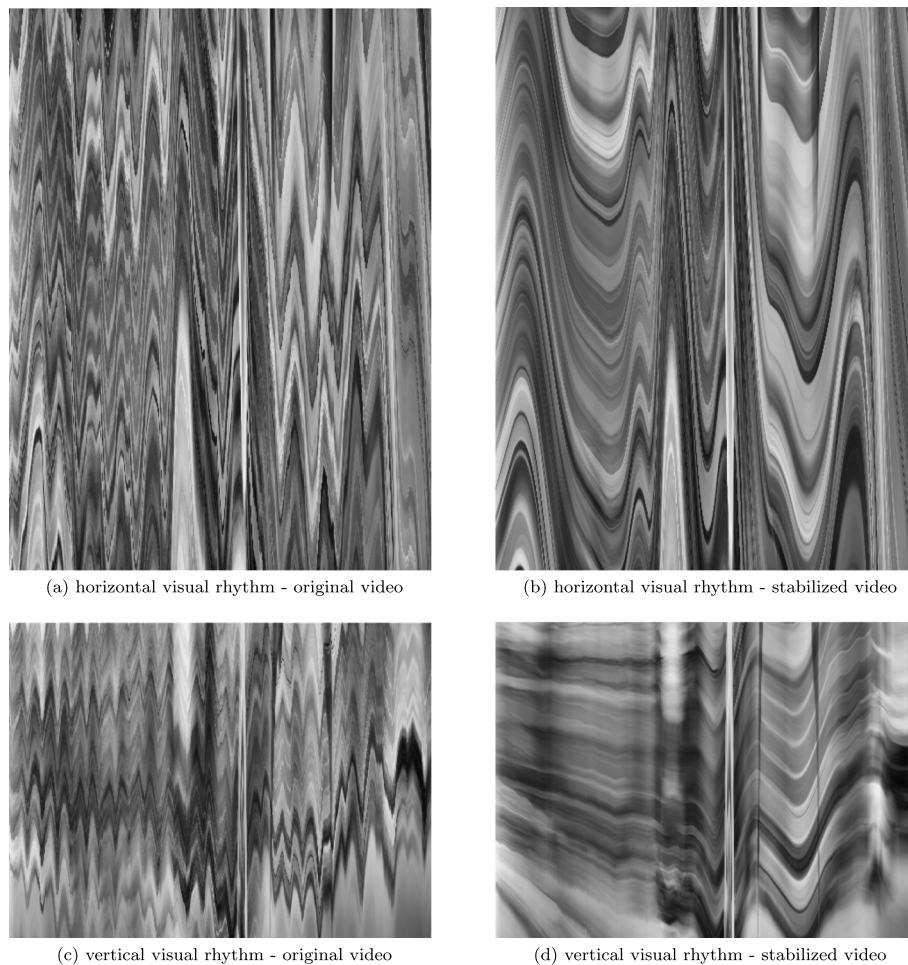
Table 2 Categories and amount of videos present in the second dataset, proposed by Liu et al. [30]

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

From the horizontal visual rhythm of the unstable video, shown in Fig. 10c, we can notice the twitches and irregularities present in the lines. On the other hand, in the horizontal visual rhythm of the stabilized video, shown in Fig. 10b, there are more continuous, well-defined and softer lines. Analogously, the vertical visual rhythm of the unstable video, shown in Fig. 10c, has twitches and irregularities that are eliminated in the visual rhythm of the stabilized video, shown in Fig. 10d. We can also observe that vertical and

(a) horizontal visual rhythm - original video

(b) horizontal visual rhythm - stabilized video

(c) vertical visual rhythm - original video

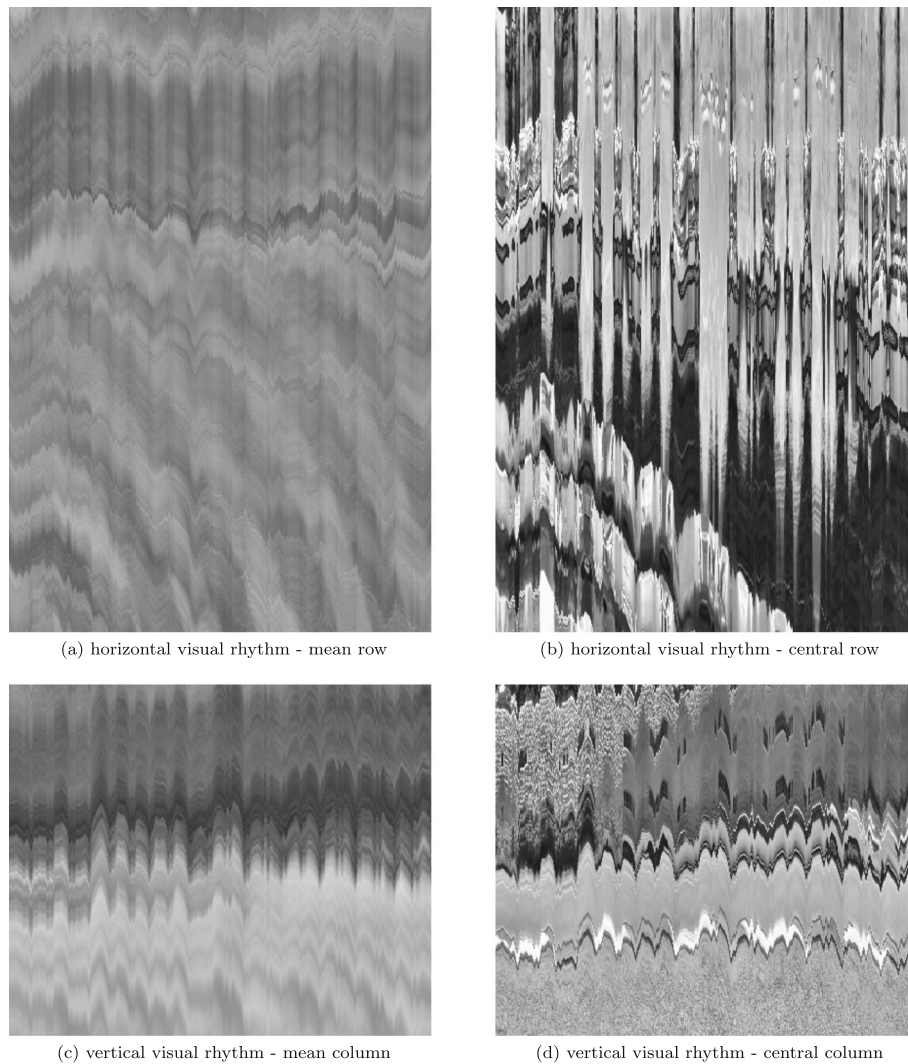(d) vertical visual rhythm - stabilized video

Fig. 10 Visual rhythms for video #12. **a** Horizontal visual rhythm for original video. **b** Horizontal visual rhythm for stabilized video. **c** Vertical visual rhythm for original video. **d** Vertical visual rhythm for stabilized video

horizontal rhythms are not influenced by each other, where certain motion regions occur in one but not in the other.

For the video Regular$_8$, we present a comparison of the visual rhythms obtained through the average of the rows or columns, and through the column or central row. In this case, we present the horizontal and vertical visual rhythms only for the unstable video.
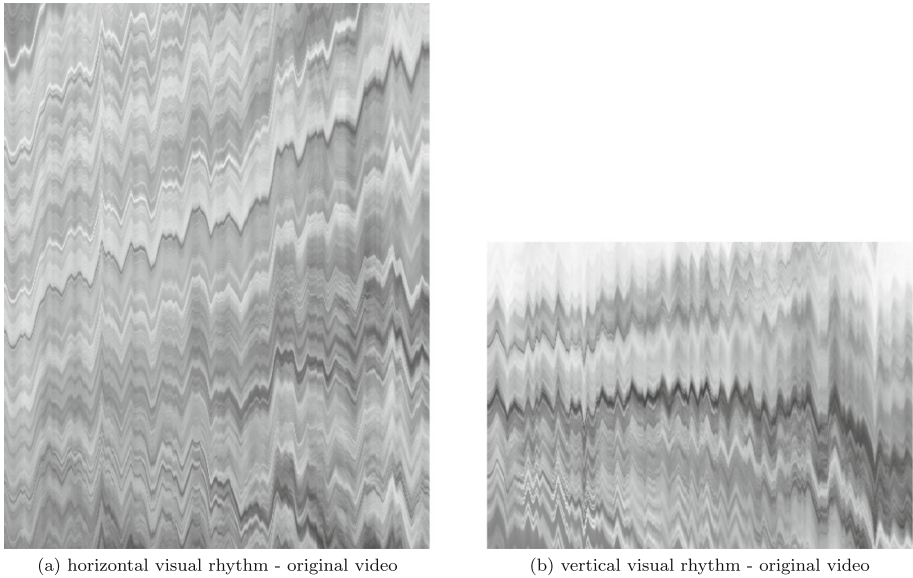
It can be seen from Fig. 11a that the visual rhythm with only one row can be negatively influenced by the vertical motion of the video, with artifacts that do not correspond to the horizontal motion, such as the discontinuities present in the rhythm, whereas the visual rhythms presented by their average are more consistent with the motion present in the video. An analogous behavior can be seen in the vertical rhythm shown in Fig. 11c.

Figure 12 presents the visual rhythms of the unstable video #1. For this video, we present the rhythms obtained after the stabilization of YouTube, in addition to a

(a) horizontal visual rhythm - mean row

(b) horizontal visual rhythm - central row

(c) vertical visual rhythm - mean column
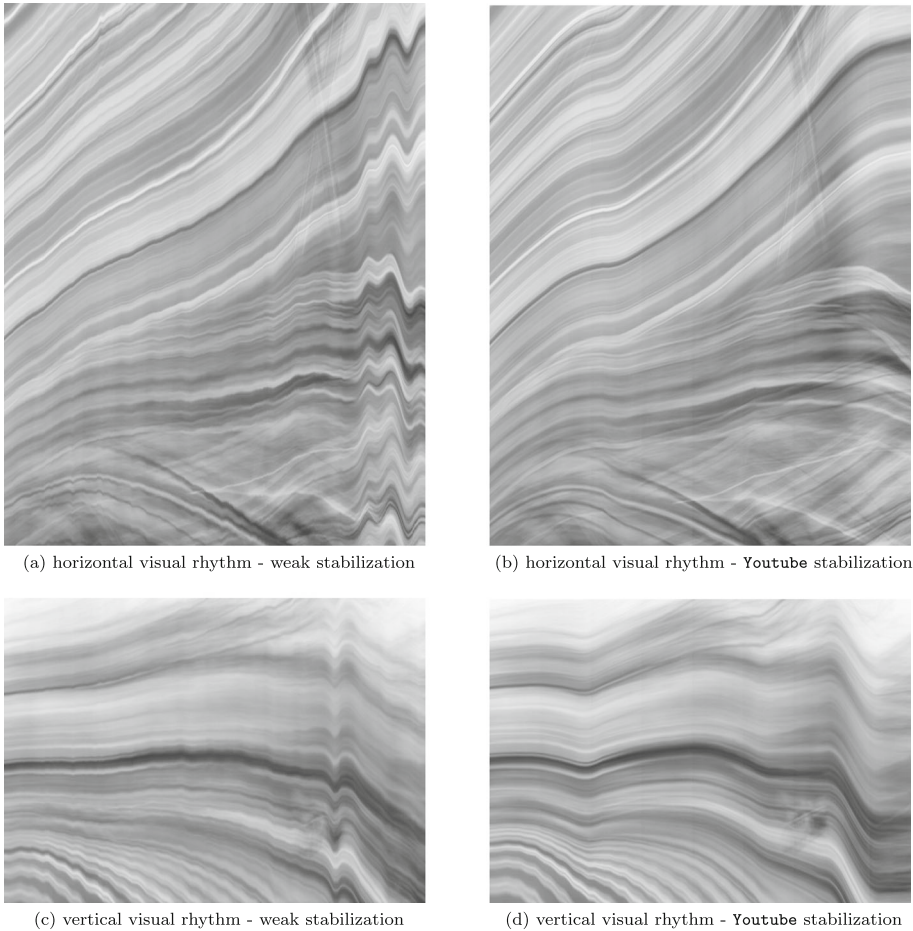
(d) vertical visual rhythm - central column

Fig. 11 Visual rhythms for original video Regular$_8$. **a** Horizontal visual rhythm with mean row. **b** Horizontal visual rhythm with central row. **c** Vertical visual rhythm with mean column. **d** Vertical visual rhythm with central column

(a) horizontal visual rhythm - original video   (b) vertical visual rhythm - original video

Fig. 12 Visual rhythms for original video #1. **a** Horizontal visual rhythm. **b** Vertical visual rhythm

(a) horizontal visual rhythm - weak stabilization   (b) horizontal visual rhythm - Youtube stabilization

(c) vertical visual rhythm - weak stabilization   (d) vertical visual rhythm - Youtube stabilization
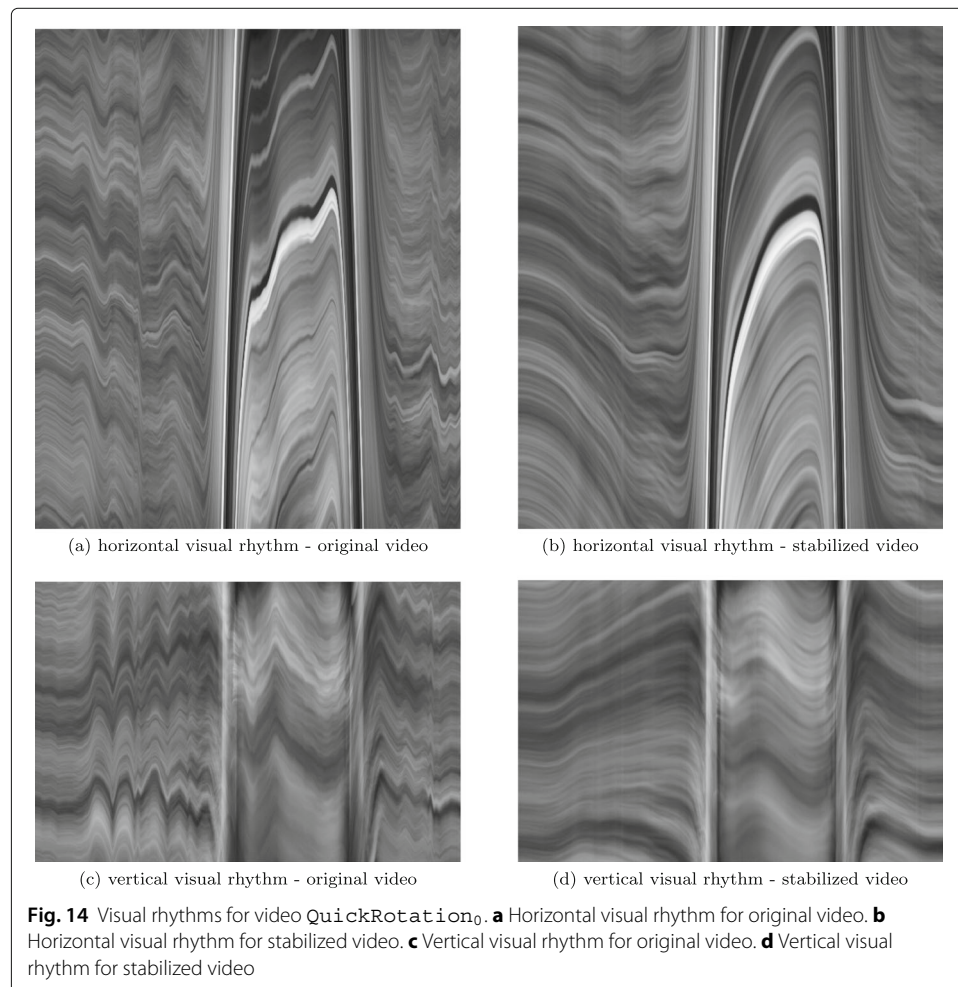
Fig. 13 Visual rhythms for stabilized video #1. **a** Horizontal visual rhythms for weak stabilization. **b** Horizontal visual rhythms for Youtube stabilization. **c** Vertical visual rhythms for weak stabilization. **d** Vertical visual rhythms for Youtube stabilization

stabilization with inferior performance. Figure 13 shows the horizontal and vertical rhythms for both versions of the stabilized video.

By comparing the visual rhythms for the unstable video and the rhythms for the stabilized videos, it is possible to confirm the validity of using visual rhythms to compare versions of stable and unstable videos. In addition, from the visual rhythms of the two different methods, illustrated in Fig. 13, we can observe the occurrence of less twitches and smoother lines throughout the entire rhythm, both for the horizontal and vertical rhythm. This shows that the visual rhythm can be used in the comparison of two different video stabilization methods.

The horizontal and vertical rhythms for the original and stabilized video $QuickRotation_0$ are shown in Fig. 14. In this case, the video was stabilized with the method proposed by Liu et al. [30]. The version of the video $QuickRotation_0$ stabilized with Youtube was not shown here since the method modified its frame rate, reducing the number of frames and making the visualization of the stabilized video considerably smaller than the original video.

Besides confirming the smoother lines obtained in the visual rhythm for the stabilized video, it is possible to observe totally vertical lines in the horizontal visual rhythms, which indicates a very fast horizontal movement of the camera. It is also possible to see that the

(a) horizontal visual rhythm - original video

(b) horizontal visual rhythm - stabilized video

(c) vertical visual rhythm - original video

(d) vertical visual rhythm - stabilized video

Fig. 14 Visual rhythms for video $QuickRotation_0$. **a** Horizontal visual rhythm for original video. **b** Horizontal visual rhythm for stabilized video. **c** Vertical visual rhythm for original video. **d** Vertical visual rhythm for stabilized video
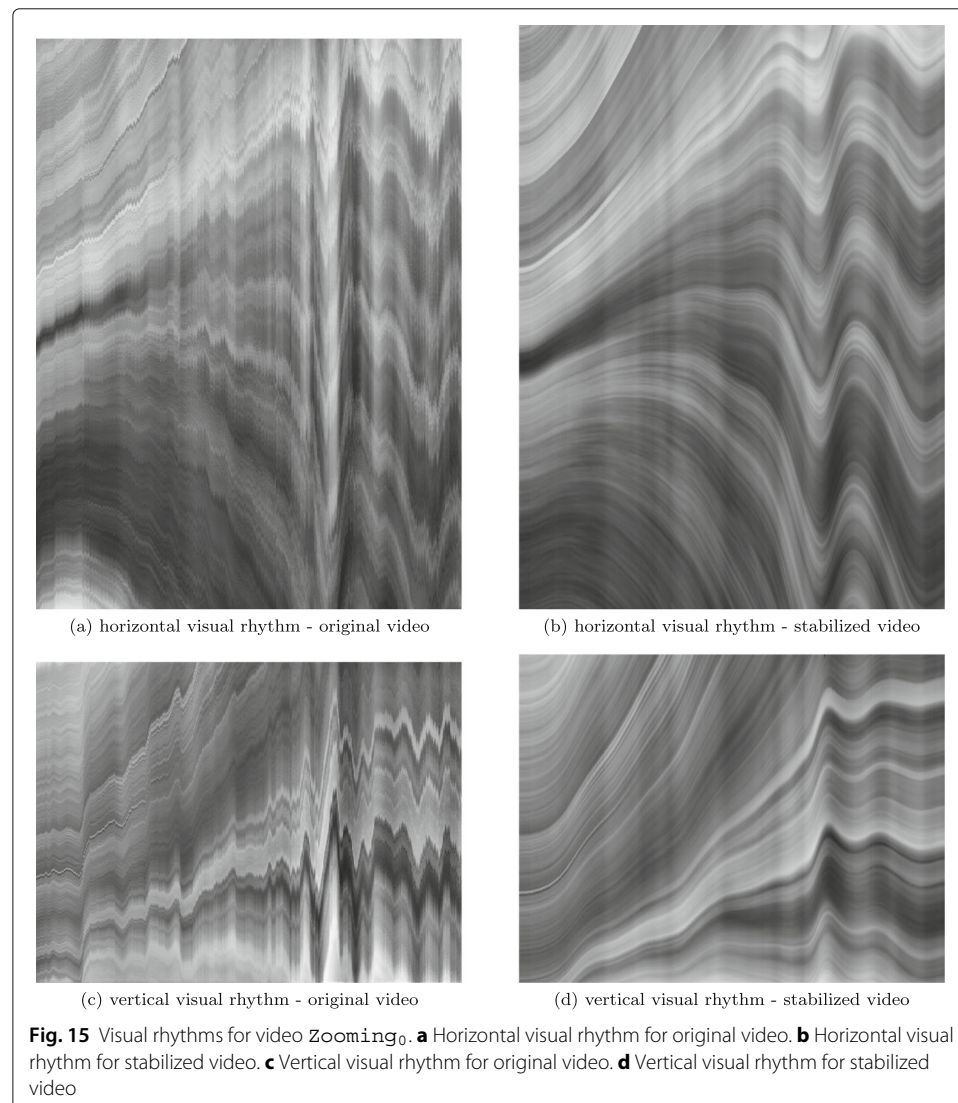
horizontal lines are inclined in their origin, which indicates that the displacement is from left to right.
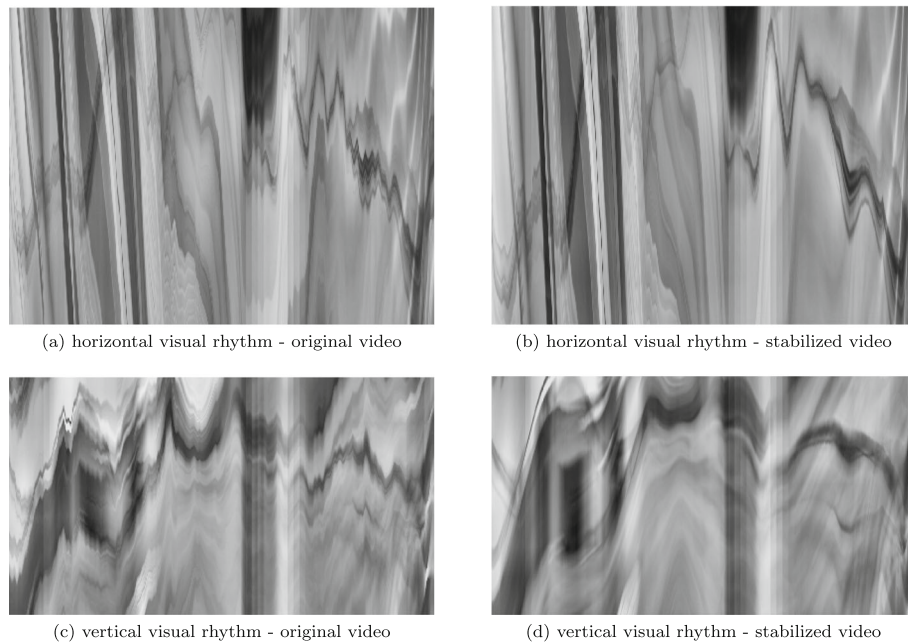
In Fig. 15, we present the horizontal and vertical visual rhythms for the original and stabilized video $Zooming_0$. The video was stabilized through the method proposed by Liu et al. [30].

In the visual rhythms for video $Zooming_0$, it is also possible to see the presence of well defined, regular lines in the visual rhythm of the stabilized video. In addition, it is possible to observe inclined and declined lines in the horizontal visual rhythms present simultaneously in the beginning of the video, which indicates the existence of zoom.

Figure 16 shows visual rhythms for a video where there is a low-texture background and a moving object. This scenario can be challenging for the proposed representation, since we do not separate the background from the objects in the construction of the rhythm. Nevertheless, the visual rhythym representation makes it possible to distinguish the unstable from the stable videos.

(a) horizontal visual rhythm - original video

(b) horizontal visual rhythm - stabilized video

(c) vertical visual rhythm - original video

(d) vertical visual rhythm - stabilized video

Fig. 15 Visual rhythms for video $Zooming_0$. **a** Horizontal visual rhythm for original video. **b** Horizontal visual rhythm for stabilized video. **c** Vertical visual rhythm for original video. **d** Vertical visual rhythm for stabilized video

(a) horizontal visual rhythm - original video   (b) horizontal visual rhythm - stabilized video

(c) vertical visual rhythm - original video   (d) vertical visual rhythm - stabilized video

Fig. 16 Visual rhythms for video with moving object on low-texture background. **a** Horizontal visual rhythm for original video. **b** Horizontal visual rhythm for stabilized video. **c** Vertical visual rhythm for original video. **d** Vertical visual rhythm for stabilized video

Table 3 reports the results of the homogeneity extracted from the horizontal and vertical visual rhythms for the video sequences listed in Table 1, where the original videos are stabilized by the YouTube method [55]. We can observe that the obtained results are able to distinguish original and stabilized videos. However, further investigation is needed regarding the extraction of other features from the co-occurrence matrix, which may be complementary to the homogeneity information. In addition, the results from the proposed metrics will be compared to objective metrics available in the literature.

Table 3 Results of homogeneity for video sequences

# Video	Original		YouTube	
	Horizontal	**Vertical**	**Horizontal**	**Vertical**
1	0.446	0.449	0.737	0.773
2	0.404	0.478	0.685	0.741
3	0.392	0.473	0.647	0.708
4	0.409	0.425	0.665	0.742
5	0.511	0.575	0.715	0.649
6	0.481	0.627	0.678	0.734
7	0.423	0.446	0.652	0.615
8	0.409	0.508	0.697	0.672
9	0.499	0.523	0.549	0.625
10	0.387	0.441	0.588	0.620
11	0.609	0.557	0.743	0.636
12	0.515	0.474	0.535	0.678
13	0.499	0.441	0.665	0.614
14	0.479	0.618	0.521	0.750
Average	0.461	0.502	0.648	0.682

5   Conclusions and future work

This work presented the use of visual rhythms for the subjective evaluation of video stabilization. The vertical visual rhythm is constructed from the average of the columns of each frame, whereas the horizontal visual rhythm is constructed from the average of the rows of each frame.

We were able to characterize and separate the horizontal and vertical movements of the video, determining how and when they happen. The stability of a video can be determined from the regularity and smoothness of the curves of each visual rhythm. In addition, the presence of more complex movements, such as zoom, can be verified in the visual rhythm.

As directions for future work, we intend to thoroughly investigate objective evaluation metrics for the stabilization of videos, calculated from the visual rhythms.

Abbreviations
2D: Two-dimensional; 3D: Three-dimensional; AVC: Advanced video coding; CLAHE: Contrast-limited adaptive histogram equalization; GLCM: Gray level co-occurrence matrix; HSV: Hue saturation value; ITF: Interframe transformation fidelity; PSNR: Peak signal-to-noise ratio; SIFT: Scale-invariant feature transform; SSIM: Structural similarity; VR: Virtual rhythm

Authors' contributions
HP and MRS contributed equally to this work. Both authors carried out the in-depth analysis of the experimental results and checked the correctness of the evaluation. Both authors took part in the writing and proof reading of the final version of the paper. The authors read and approved the final manuscript.

Availability of data and materials
Data are publicly available.

Competing interests
The authors declare that they have no competing interests.

References
1. A. A. Amanatiadis, I. Andreadis, Digital image stabilization by independent component analysis. IEEE Trans. Instrum. Meas. **59**(7), 1755–1763 (2010)
2. J. Y. Chang, W. F. Hu, M. H. Cheng, B. S. Chang, Digital image translational and rotational motion stabilization using optical flow technique. IEEE Trans. Consum. Electron. **48**(1), 108–115 (2002)
3. S. Ertürk, Real-time digital image stabilization using Kalman filters. Real Time Imaging. **8**(4), 317–328 (2002)
4. R. Jia, H. Zhang, L. Wang, J. Li, in *International Conference on Artificial Intelligence and Computational Intelligence. vol. 3*, Digital image stabilization based on phase correlation (IEEE, 2009). https://doi.org/10.1109/aici.2009.489
5. S. J. Ko, S. H. Lee, K. H. Lee, Digital image stabilizing algorithms based on bit-plane matching. IEEE Trans. Consum. Electron. **44**(3), 617–622 (1998)
6. S. Kumar, H. Azartash, M. Biswas, T. Nguyen, 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)
7. C. T. Lin, C. T. Hong, C. T. Yang, Real-time digital image stabilization system using modified proportional integrated controller. IEEE Trans. Circ. Syst. Video Technol. **19**(3), 427–431 (2009)
8. L. Marcenaro, G. Vernazza, C. S. Regazzoni, in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, Image stabilization algorithms for video-surveillance applications (IEEE, 2001). https://doi.org/10.1109/icip.2001.959025
9. C. Morimoto, R. Chellappa, in *13th International Conference on Pattern Recognition. vol. 3*, Fast electronic digital image stabilization (IEEE, 1996). https://doi.org/10.1109/icpr.1996.546956
10. Y. G. Ryu, M. J. Chung, Robust online digital image stabilization based on point-feature trajectory without accumulative global motion estimation. IEEE Sig. Process. Lett. **19**(4), 223–226 (2012)
11. J. Li, T. Xu, K. Zhang, Real-time feature-based video stabilization on FPGA. IEEE Trans. Circ. Syst. Video Technol. **27**(4), 907–919 (2017)
12. M. Okade, G. Patel, P. K. Biswas, Robust learning-based camera motion characterization scheme with applications to video stabilization. IEEE Trans. Circ. Syst. Video Technol. **26**(3), 453–466 (2016)
13. M. R. Souza, H. Pedrini, Combination of local feature detection methods for digital video stabilization. SIViP. **12**(8), 1513–1521 (2018)

14. M. R. Souza, H. Pedrini, Digital video stabilization based on adaptive camera trajectory smoothing. EURASIP J. Image Video Process. **2018**(37), 1–11 (2018)

15. M. R. Souza, L. F. R. Fonseca, H. Pedrini, Improvement of global motion estimation in two-dimensional digital video stabilization methods. IET Image Process. **12**(12), 2204–2211 (2018)

16. M. V. M. Cirne, H. Pedrini, in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, A video summarization method based on spectral clustering (Springer, 2013), pp. 479–486. https://doi.org/10.1007/978-3-642-41827-3_60

17. M. V. M. Cirne, H. Pedrini, in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Summarization of videos by image quality assessment (Springer, 2014), pp. 901–908

18. T. S. Huang, *Image Sequence Analysis. vol. 5*. (Springer Science & Business Media, Berlin, 2013)

19. B. Cardani, Optical image stabilization for digital cameras. IEEE Control Syst. **26**(2), 21–22 (2006)

20. C. Buehler, M. Bosse, L. McMillan, in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Non-metric image-based rendering for video stabilization (IEEE, 2001). https://doi.org/10.1109/cvpr.2001.991019

21. G. Zhang, W. Hua, X. Qin, Y. Shao, H. Bao, Video stabilization based on a 3D perspective camera model. Vis. Comput. **25**(11), 997–1008 (2009)

22. R. C. Gonzalez, R. E. Woods, *Digital Image Processing*. (Prentice Hall, Upper Saddle River, 2002)

23. M. Niskanen, O. Silvén, M. Tico, in *IEEE International Conference on Multimedia and Expo*, video stabilization performance assessment (IEEE, 2006). https://doi.org/10.1109/icme.2006.262522

24. S. Battiato, G. Gallo, G. Puglisi, S. Scellato, in *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, SIFT features tracking for video stabilization (IEEE, 2007). https://doi.org/10.1109/iciap.2007.4362878

25. S. Choi, T. Kim, W. Yu, in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Robust video stabilization to outlier motion using adaptive RANSAC (IEEE, 2009). https://doi.org/10.1109/iros.2009.5354240

26. G. Puglisi, S. Battiato, A robust image alignment algorithm for video stabilization purposes. IEEE Trans. Circ. Syst. Video Technol. **21**(10), 1390–1400 (2011)

27. D. Shukla, R. K. Jha, A robust video stabilization technique using integral frame projection warping. SIViP. **9**(6), 1287–1297 (2015)

28. B. H. Chen, A. Kopylov, S. C. Huang, O. Seredin, R. Karpov, S. Y. Kuo, et al., Improved global motion estimation via motion vector clustering for video stabilization. Eng. Appl. Artif. Intell. **54**, 39–48 (2016)

29. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

30. S. Liu, L. Yuan, P. Tan, J. Sun, Bundled camera paths for video stabilization. ACM Trans. Graph. **32**(4), 78 (2013)

31. H. Qu, L. Song, G. Xue, in *2013 Visual Communications and Image Processing (VCIP)*, Shaking video synthesis for video stabilization performance assessment (IEEE, 2013). https://doi.org/10.1109/vcip.2013.6706422

32. D. G. Lowe, in *Object Recognition from Local Scale-Invariant Features*, Object recognition from local scale-invariant features (IEEE, 1999). https://doi.org/10.1109/iccv.1999.790410

33. B. Chen, J. Zhao, Y. Wang, in *Proceedings of the 2016 International Conference on Advanced Materials Science and Environmental Engineering*, Research on evaluation method of video stabilization (Atlantis Press, 2016). https://doi.org/10.2991/amsee-16.2016.67

34. K. Ratakonda, in *ISCAS '98. Proceedings of the 1998 IEEE International Symposium on Circuits and Systems (Cat. No.98CH36187)*, Real-time digital video stabilization for multi-media applications (IEEE, 1998). https://doi.org/10.1109/iscas.1998.698760

35. A. Litvin, J. Konrad, W. C. Karl, in *Proceedings of SPI. vol. 5022*, Probabilistic video stabilization using Kalman filtering and mosaicing (International Society for Optics and Photonics, 2003), pp. 663–674

36. H. C. Chang, S. H. Lai, K. R. Lu, in *2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763)*, A robust and efficient video stabilization algorithm (IEEE, 2004). https://doi.org/10.1109/icme.2004.1394117

37. Y. Matsushita, E. Ofek, W. Ge, X. Tang, H. Y. Shum, Full-frame video stabilization with motion in painting. IEEE Trans. Pattern Anal. Mach. Intell. **28**(7), 1150–1163 (2006)

38. B. Y. Chen, K. Y. Lee, W. T. Huang, J. S. Lin, Wiley Online Library. Capturing intention-based full-frame video stabilization. Comput. Graph. Forum. **27**(7), 1805–1814 (2008)

39. Y. Shen, P. Guturu, T. Damarla, B. P. Buckles, K. R. Namuduri, Video stabilization using principal component analysis and scale invariant feature transform in particle filter framework. IEEE Trans. Consum. Electron. **55**(3), 1714–1721 (2009)

40. J. Yang, D. Schonfeld, M. Mohamed, Robust video stabilization based on particle filter tracking of projected camera motion. IEEE Trans. Circ. Syst. Video Technol. **19**(7), 945–954 (2009)

41. N. Joshi, W. Kienzle, M. Toelle, M. Uyttendaele, M. F. Cohen, Real-time hyperlapse creation via optimal frame selection. ACM Trans. Graph. **34**(4), 63 (2015)

42. Q. Zheng, M. Yang, A video stabilization method based on inter-frame image matching score. Glob. J. Comput. Sci. Technol. **17**(1-F) (2017)

43. R. Borgo, M. Chen, B. Daubney, E. Grundy, G. Heidemann, B. Höferlin, et al., Wiley Online Library. State of the art report on video-based graphics and video visualization. Comput. Graph. Forum. **31**(8), 2450–2477 (2012)

44. K. Schoeffmann, M. Lux, M. Taschwer, L. Boeszoermenyi, in *2009 IEEE International Conference on Multimedia and Expo*, Visualization of video motion in context of video browsing (IEEE, 2009). https://doi.org/10.1109/icme.2009.5202582

45. M. G. Chung, J. Lee, H. Kim, S. M. H. Song, W. M. Kim, Automatic video segmentation based on spatio-temporal features. Korea Telecom J. **4**(1), 4–14 (1999)

46. F. B. Valio, H. Pedrini, N. J. Leite, in *16th Iberoamerican Congress on Pattern Recognition*, Fast rotation-invariant video caption detection based on visual rhythm (Pucón, Chile, 2011), pp. 157–164

47. A. Pinto, W. R. Schwartz, H. Pedrini, A. Rezende Rocha, Using visual rhythms for detecting video-based facial spoof attacks. IEEE Trans. Inf. Forensic. Secur. **10**(5), 1025–1038 (2015)

48. B. S. Torres, H. Pedrini, Detection of complex video events through visual rhythm. Vis. Comput. **34**(2), 145–165 (2018)

49. A. Silva Pinto, H. Pedrini, W. Schwartz, A. Rocha, in *25th Conference on Graphics, Patterns and Images Ouro Preto-MG*, Video-based face spoofing detection through visual rhythm analysis (IEEE, Brazil, 2012), pp. 221–228

50. T. P. Moreira, D. Menotti, H. Pedrini, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, First-person action recognition through visual rhythm texture description, (New Orleans, LA, USA, 2017), pp. 2627–2631. https://doi.org/10.1109/icassp.2017.7952632

51. K. Zuiderveld, in *Graphics Gems IV*, Contrast limited adaptive histogram equalization (Academic Press Professional, Inc., 1994), pp. 474–485. https://doi.org/10.1016/b978-0-12-336156-1.50061-6

52. I. Sobel, G. Feldman, A 3x3 isotropic gradient operator for image processing. Talk Stanf. Artif. Proj, 271–2 (1968)

53. N. Otsu, A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man. Cybernet. **9**(1), 62–66 (1979)

54. R. M. Haralick, K. Shanmugam, Textural features for image classification. IEEE Trans. Syst. Man. Cybernet. **SMC-3**(6), 610–621 (1973)

55. M. Grundmann, V. Kwatra, I. Essa, in *IEEE Conference on Computer Vision and Pattern Recognition*, Auto-directed video stabilization with robust L1 optimal camera paths (IEEE, 2011), pp. 225–232. https://doi.org/10.1109/cvpr.2011.5995525

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.