

Digital Video Stabilization: Algorithms and Evaluation

Marcos Roberto e Souza*
Institute of Computing
University of Campinas
Campinas, Brazil – 13083-852
Email: marcoosrs@gmail.com

Helio Pedrini
Institute of Computing
University of Campinas
Campinas, Brazil – 13083-852
Email: helio@ic.unicamp.br

Abstract—Several devices have allowed the acquisition and editing of videos in various circumstances, such as digital cameras, smartphones and other mobile devices. However, the use of cameras under adverse conditions usually results in non-precise motion and occurrence of shaking, which may compromise the stability of the obtained videos. To overcome such problem, digital stabilization aims to correct camera motion oscillations that occur in the acquisition process, particularly when the cameras are mobile and handled in adverse conditions, through software techniques - without the use of specific hardware - to enhance visual quality either with the intention of enhancing human perception or improving final applications, such as detection and tracking of objects. This is important in order to avoid hardware cost and indispensable for videos already recorded. This work proposed three methods to perform digital video stabilization and two other techniques to evaluate video stabilization quality.

I. INTRODUCTION

In this work, we are particularly interested in investigating two-dimensional (2D) video stabilization methods [1]–[18], in which geometric transformations are employed to represent frame-to-frame motion and stabilize the videos. The reason for this interest is that even though three-dimensional (3D) methods allow higher quality stabilization, 2D methods have a lower computational cost and are more robust to a variety of situations, which causes them to be constantly preferred in practice. The 2D digital video stabilization process is usually divided into three main steps: (i) camera motion estimation, where the motions performed by the camera are estimated, constructing a path that corresponds to the one traveled by the camera, (ii) removal of unwanted motion, which smooths the unstable video motion, and (iii) generation of the corrected video, which transforms the video frames according to the remaining motion.

This work aimed to investigate and evaluate digital video stabilization methods for correcting disturbances and instabilities that occur during the process of video capture. It also proposed novel methods for digital video stabilization and for qualitative evaluation of the video stabilization process. Experiments were conducted on several video sequences. A comparative analysis of the results obtained with the proposed

method and with other approaches of the literature were presented and discussed.

The main contributions of this work are: (i) a consensual approach to combining different methods of local features in motion estimation. We experimentally demonstrated that the results of individual methods could be improved by combining different methods; (ii) an approach that detected failures in the global motion estimation obtained through local features and proposed an optimization technique to calculate a new estimate of the corrected motion. Experiments showed that estimation of the optimization method is considerably superior when compared to the individual use of local features. The state-of-the-art stabilization method used in YouTube [19] was also used for comparison, which presented typical flaws when using local features to motion estimation, obtaining in these cases a worse result than the method proposed; (iii) a new technique for removing unwanted motion based on the Gaussian filter to smooth the camera path. Experiments demonstrated the effectiveness of the method, which generated videos with proper stabilization rate while maintaining a reasonable amount of frame pixels; (iv) new techniques for the qualitative evaluation of video stabilization through visual representations based on visual rhythms and motion energy image. We proposed a visualization scheme based on visual rhythms to represent the behavior of the motion present in a video. In addition, a visualization based on motion energy image was used to represent the amount of motion present in a video. Both proposed evaluation approaches were intended for human beings to assess the quality of the stabilization. Experimental results demonstrated that the both visual representations were effective to evaluate the stability of camera motion by differentiating stable and unstable videos.

II. VIDEO STABILIZATION METHODS

In this section, we present the proposed methods for digital stabilization of videos [10], [11], [20]–[32]. Initially, the camera motion is estimated. In this step, we propose two approaches: (i) the combination of local features, which fuses different local feature methods in a consensual way in order to estimate the global camera motion, and (ii) an optimization strategy based on the structural similarity index that considers spatial and temporal information to correct cases of failure

* This is a summary of an M.Sc. dissertation.

in the global motion estimation. A smoother motion is then calculated from the camera motion. For this, we propose an adaptive Gaussian filter, in which the intensity of the filter is changed adaptively along the video according to the behavior of the estimated motion. Subsequently, the frames of the video are transformed in order to follow the smoothed motion. In addition, frame borders are cropped to keep only the useful information. Finally, the final video is submitted to a qualitative assessment. In the evaluation stage, we propose two approaches: (i) evaluation of the behavior of motion present in the video based on visual rhythms, and (ii) evaluation of the amount of motion present in the video based the motion energy image. A detailed description of all the methods briefly presented in this section can be found in Chapter 3 of the Master’s Dissertation [33].

A. Motion Estimation with Local Combined Features

Initially, we consider a set of methods M_f to be combined. For each method $m \in M_f$, its respective detection and description are applied for each pair of frames. Then, the local features are matched considering the local features of each method m separately. The correspondence is performed using the brute-force method with cross-checking [34]. Given the sets of local features that belong to the adjacent frames f_t and f_{t+1} , such that $x_i^m \in f_t$ and $y_j^m \in f_{t+1}$, we calculate the Euclidean distance (or Hamming distance when the descriptor is binary) between the feature vectors for each pair of local features x_i^m and y_j^m . Thus, x_i^m corresponds to y_j^m if and only if x_i^m is the closest local feature to y_j^m , and y_j^m the closest to x_i^m . Before combining the matches of each method, a pre-evaluation is performed based on harmonic mean of the quadratic errors to discard those that would perform poorly or could contribute negatively to the final combination. Then, a consensual combination is applied in the remaining methods, such that only local features that are consistent with their transformation are considered as final local features.

Let the sets of local features be $F_{all} = F_0 \cup F_1 \dots \cup \dots F_i$, and $F'_{all} = F'_0 \cup F'_1 \dots \cup \dots F'_i$, where F_i are the local features of the frame f_t , F'_i the features f_{t+1} , and i the methods that passed in the previous step. The transformation matrix of F_{all} to F'_{all} , called the global transformation matrix, is estimated. Next, a transformation matrix is calculated for each possible combination of methods. Each previously calculated transformation matrix H^c is applied to the set of local features F_{all} , obtaining F''_{all} . For each local feature of F''_{all} , the quadratic error with its equivalent local feature in F'_{all} is calculated. If this quadratic error is less than or equal to the mean square error of the global transformation matrix, such local feature is considered inlier of H^c , and otherwise, outlier of H^c . Finally, the combined local features are taken as the inliers of the transformation matrix that has the largest number of inliers.

B. Motion Estimation with Spatio-Temporal Optimization

We applied a consistency check on the matrix estimated by the local features, comparing it with the estimated (and final) in the previous frame pair. As premise, we consider that the

previous frames have a correct motion estimation. Moreover, difference in the camera motion and by the objects between two frames is small. Even though sudden movements may occur, the video sequences have typically several frames per second, which makes the difference between adjacent frames more gradual.

We calculate the structural similarity index (SSIM) [35] image S_t for the frame pair f'_t and f_{t+1} . Then, we calculate the absolute difference of the image pixels of S_t , with the image S_{t-1} . Thus, we obtain the difference image $D_t = S_t - S_{t-1}$, which indicates how similar the image of similarity is in relation to the previous image. Since the images S_t and S_{t-1} consider the transformed frames, the regions of them with high values indicate the presence of remaining movements. If the estimation is correct, such motion corresponds to objects, not to the camera. Thus, if the movements are similar and spatially close in both images (S_t and S_{t+1}), D_t tends to have lower values.

For an estimate to be considered potentially inconsistent, at least one of two inequalities must be satisfied. First inequality checks the variation between (i) the similarity between the reference frame f_{t+1} and the transformed frame f'_t and (ii) the similarity obtained in the previous frame pairs. The other inequality checks the variation between (i) the difference of the similarity of the current and previous frame and (ii) the difference between the pairs of previous frames.

After detecting an inconsistency, the new motion estimation is calculated. This new estimate refers to the transformation matrix that minimizes an objective function based on the similarity value and the difference value. This objective function is based on

$$f = \left\{ (\alpha)(1 - \text{mean}(S_{new})) + (1 - \alpha)(\text{mean}(D_{new})) \right\} \quad (1)$$

where S_{new} is the SSIM image between the reference frame f_{t+1} and the frame transformed by the matrix being minimized f'_t , considering the cropping area obtained by the same matrix, whereas D_{new} is the difference image between S_{new} and S_{t-1} . These values are considered only if three conditions are met: cropping condition, coherence condition, and boundary condition. In this work, the Powell method [36] is used to minimize the objective function presented in Equation 1.

C. Adaptive Gaussian Filter

After estimating the final similarity matrices for each pair of adjacent frames of the video, a trajectory is calculated for each of the factors. In this work, we consider a vertical translation factor, a horizontal translation factor, a rotation factor and a scaling factor. Each factor f of the matrix is decomposed and the trajectory of each of them is calculated in order to accumulate its previous values, expressed as

$$t_i^f = t_{i-1}^f + \Delta_i^f \quad (2)$$

where t_i is the value of a given trajectory in the i -th position, and Δ_i^f is the value of factor f for the i -th similarity matrix previously estimated. The trajectories are then smoothed. The equations presented in the remainder of the text will be always

applied to the trajectories of each factor separately. Thus, the factor index f will be omitted in order to not overload the notation.

In our method, σ is computed in such a way that it has smaller values in regions with intense motion. Thus, the trajectory will be smoothed by considering a distinct value for σ_i at each point i . To determine the value of σ_i , a sliding window of size twice as large as the frame-rate measure is applied, so that the window information lasts for two video seconds. The ratio r_i is expressed as

$$r_i = \left(1 - \frac{\mu_i}{\text{max_value}}\right)^2 \quad (3)$$

where max_value corresponds to either width in the horizontal translation trajectory or height in the vertical translation trajectory. Value μ_i is calculated in such a way to give higher weights to points closer to i , where μ_i is expressed as

$$\mu_i = \frac{\sum_{j \in W_i, j \neq i} G(|j - i|, \sigma_\mu) \Delta_j}{\sum_{j \in W_i} G(|j - i|, \sigma_\mu)} \quad (4)$$

where j is the index of each point in the window of i , whereas $G()$ is a Gaussian function with σ calculated as

$$\sigma_\mu = (\text{FPS})(1 - \text{CV}) \quad (5)$$

where FPS is the video frames per second, and CV is the coefficient of variation of the absolute values of the trajectory that are inside the window. Since the value of CV lies between 0 and 1, its final value is limited to 0.9 in order for σ_μ not to have null values. Therefore, σ_μ makes the actual size of the window adaptive, such that the higher the variation of motion inside the window, the higher the weight given to the central points.

Assuming that r_i ranges between 0 and 1, a linear transformation is applied to obtain a proper interval for the Gaussian filter. An exponential transformation is then applied to σ_i values to amplify their magnitude. After calculating σ_i for each point of the trajectory, its values are lightly smoothed by a Gaussian filter with $\sigma = 5$, chosen empirically. Finally, the Gaussian filter is applied n times (once for each point in the trajectory), generating a smoothed trajectory (indexed by k) for each σ_i previously calculated. The final smoothed trajectory corresponds to the concatenation of points for each of the generated trajectories, and the k -th trajectory contributes with its k -th point. Thus, an adaptive smoothed path is obtained.

D. Evaluation of Stabilization based on Visual Rhythms

In the evaluation based on visual rhythms, two different path directions are considered: horizontal and vertical. The vertical rhythm extracts the information from the columns of each frame, while the horizontal rhythm takes the information from the lines of each frame. For both path directions, the rhythm is obtained from the sequential concatenation of the information, so that the j -th column of the visual rhythm image corresponds to the information in the j -th frame. The width of a visual rhythm corresponds to the number of frames of the video,

whereas its height corresponds to the height or width of the frames for the vertical or horizontal rhythm, respectively.

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. Thus, the average of the columns or rows is proposed 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. As post-processing, we apply an adaptive histogram equalization technique through the Contrast Limited Adaptive Histogram Equalization (CLAHE) [37].

E. Evaluation of Stabilization based on Motion Energy Image

We also propose a stabilization evaluation method based on the motion energy image. For each video frame i , the difference $\text{Diff}_{i,j}$ of the gray level intensities of each pixel is calculated. This is done by considering the pre-processed frames through a Gaussian filter, 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, in which 1 is assigned to the pixel with difference greater than a certain threshold, and 0 otherwise.

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

$$\text{MEI}_i = \frac{\sum_{j \in \Omega_i, j \neq i} G(|i - j|, \sigma) \text{Diff}_{i,j}}{\sum_{j \in \Omega_i, j \neq i} G(|i - j|, \sigma)} \quad (6)$$

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

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 the 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. A pseudocolor transformation is applied, so that high gray-level intensity values are mapped to red, whereas lower intensities to blue.

III. EXPERIMENTS

Results obtained in the experiments are described in this section. Three databases are used to evaluate the effectiveness of the proposed video stabilization methods. The first consists of eleven videos available in the GaTech VideoStab [19] dataset and three others collected separately. The second, available by Liu et al. [38], consists of 139 videos divided into categories. Finally, we create a dataset that is complementary to the others, in which four videos with moving objects in the foreground and with little representative backgrounds are collected separately. All the experiments conducted on the datasets, as well as the performed comparisons, are presented and discussed in detail in Chapter 4 of the Master's Dissertation [33].

A. Local Combined Features

Tables I present the results obtained, in which we can see a better combination performance for most videos for PSNR. In average values where the combination is not the best result, it tends to be close to the best. For example, in video #7 the combination has a better result than the MSER and STAR methods, whereas in video #13 the result is greater than MSER and SIFT.

These results show that the application of the combination strategy obtains a greater robustness in the motion estimation. In the second dataset, the results were also considerably better for some videos. In addition, we notice that the results obtained with the combination are lower than some individual results, for instance, with the SURF method. This occurs because the results achieved with SURF for these two datasets are already very good, correctly estimating the global motion between two frames in practically all cases, without being able to obtain better results.

Table I

PSNR FOR DIFFERENT LOCAL FEATURES IN MOTION ESTIMATION.

Video	Original	MSER	SIFT	STAR	Combination
1	18.792	29.537	30.378	29.783	30.513
2	20.390	28.198	29.863	29.300	29.993
3	16.186	23.382	24.853	24.925	25.772
4	19.965	27.984	32.678	31.377	32.985
5	23.276	32.797	33.531	32.787	33.726
6	19.680	25.279	29.893	27.136	29.988
7	24.108	26.287	29.520	25.486	28.801
8	17.880	27.356	28.115	27.806	28.249
9	19.248	23.587	25.329	23.730	25.463
10	12.971	18.560	18.937	18.885	19.016
11	21.487	27.971	28.621	27.600	28.424
12	15.081	23.281	23.761	23.279	23.922
13	23.840	27.963	26.558	28.638	28.516
14	18.064	26.168	27.205	24.731	26.987
Average	19.355	26.311	27.803	26.819	28.025

B. Spatio-Temporal Optimization

Figure 1 present a failure situation of local features for different videos, as well as the correction performed with our method. Matches considered as inliers by the RANSAC method are drawn in blue and green, whereas the outlier matches are drawn in pink and yellow.

In the case presented, we can see that the matches of the objects were considered as inliers, which made the movements of the object, not the camera, compensated. On the other hand, our optimization-based method obtained excellent results, finding the transformation matrix that matches the motion performed by the camera.

Higher values of similarity measures, such as PSNR or SSIM, may indicate a better quality in the motion estimation for most cases. However, there are cases where such measures do not indicate the correct estimate and, therefore, a simple optimization that takes the measures into account would not be efficient. Figure 2 presents different matches for the same

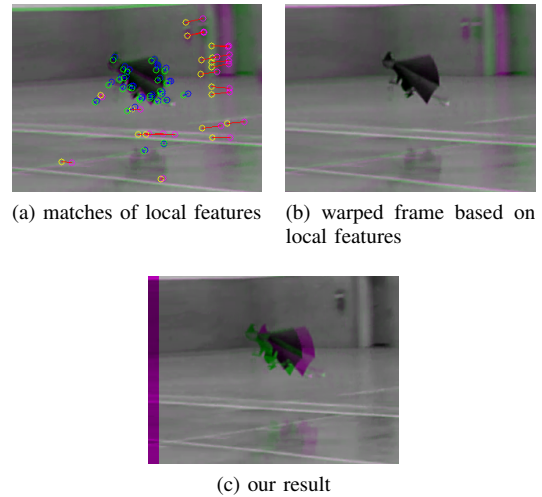


Figure 1. Motion estimation for the 129th frame of video `ours6`.

pair of frames, where different values of PSNR and SSIM are obtained. It can be observed that higher values are obtained in incorrect cases. Since background is unrepresentative, higher similarity is obtained if object matching is done. However, this is semantically incorrect since the object is in motion.

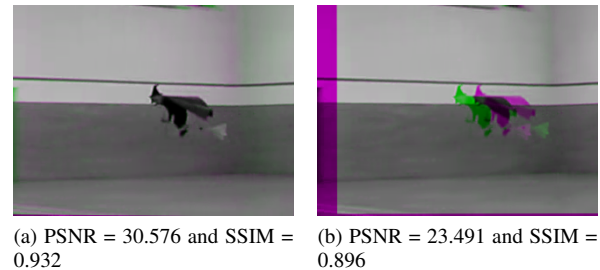


Figure 2. Different matches for the 40th frame of video `ours7`.

Figure 3 shows the vertical visual rhythms obtained both from unstable video #13 and videos stabilized with the three versions. We can verify that the beginning of the YouTube visual rhythm has vertical lines, unlike the original rhythm and the other stabilized versions. This indicates that the YouTube method has added an artificial motion to the video, which does not correspond to the desired purpose. In this case, the image remains static in the first frames, without the occurrence of motion. This occurred because the motion estimation computed by the YouTube method takes into account the movement of the object, which is consequently compensated in the stabilization.

The visual rhythms, illustrated in Figure 3, show that our method corrects several instabilities or discontinuities that occur in the estimation based on local features. This is especially noticeable at the beginning and end of the visual rhythms. Compared to the visual rhythm obtained with YouTube, the rhythm generated by the proposed method is significantly more regular, representing a better quality in the video stabilization process.

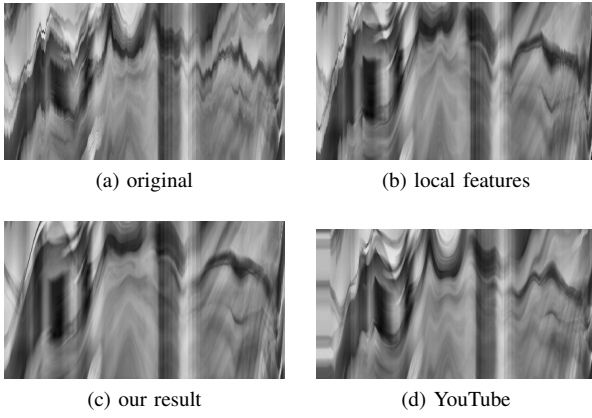


Figure 3. Vertical visual rhythms for video `ours7`.

C. Adaptive Gaussian Filter

Figures 4a, 4b and 4c show the trajectory generated by considering the horizontal translational factor (blue) and the obtained smoothing (green), respectively, using the Gaussian filter with $\sigma = 20$ and $\sigma = 40$, besides the adaptive version proposed in this work. It is possible to observe that the smoothing is applied at different degrees along the trajectory.

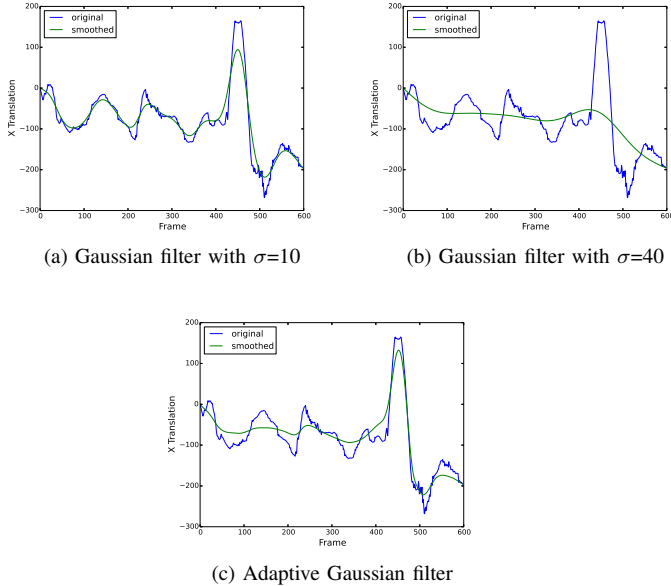


Figure 4. Smoothing of camera motion trajectories.

Table II present the results obtained with the adaptive version proposed in this work. The proposed adaptive Gaussian achieved values comparable to the original version, maintaining considerably more pixels (up to 50 percentage points more). From the results obtained in the second dataset, the gain in the percentage of pixels held was more significant in the `QuickRotation`, `Zooming` and `Running` categories. Compared to the `Youtube` method, we can observe a certain parity for both methods in terms of ITF, with a slight advantage of the `Youtube`, while the maintained pixels are in general comparable and, when lower, they do not differ much.

Table II
COMPARISON OF ITF VALUES AND HOLD PIXELS (%) BETWEEN SEMI-ADAPTIVE GAUSSIAN FILTER AND ADAPTIVE GAUSSIAN FILTER.

# Video	Original ITF	Gaussian Filter		Adaptive Gaussian	
		ITF	Hold Pixels	ITF	Hold Pixels
1	18.793	27.738	69.276	27.455	74.500
2	20.390	29.331	71.750	28.914	75.781
3	16.186	22.559	72.972	22.090	76.056
4	19.965	33.380	48.958	27.931	62.465
5	23.277	28.660	02.540	27.360	53.385
6	19.681	29.804	67.891	29.077	70.838
7	24.109	28.510	60.495	28.876	73.667
8	17.881	25.448	70.648	25.182	73.284
9	19.248	23.251	25.797	21.435	57.139
10	12.972	18.453	17.519	16.381	70.296
11	21.487	26.826	43.599	25.659	57.260
12	15.081	-	0.0000	17.895	59.847
13	23.841	30.621	70.312	29.987	71.719
14	18.065	20.265	07.448	19.773	54.146
Average	19.355	26.526	44.943	24.858	66.455

D. Visual Rhythms

Figure 5 presents the visual rhythms generated for the video #12 before and after the video stabilization process. This experiment was done to verify if an unstable video could be differentiated from a stabilized video.

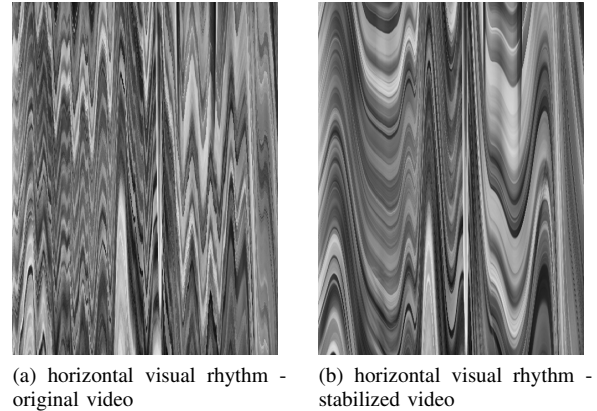
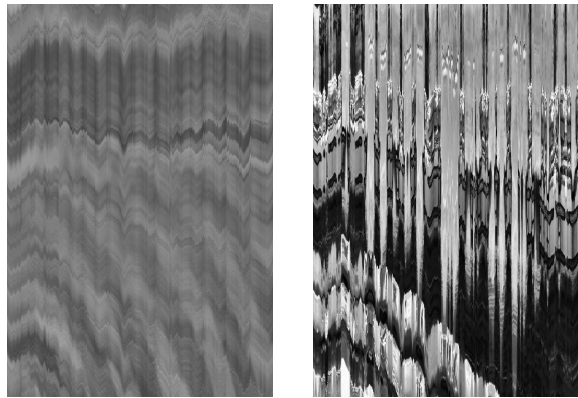


Figure 5. Visual rhythms for video #12.

From the horizontal visual rhythm of the unstable video, shown in Figure 5a, 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 Figure 5b, there are more continuous, well defined and softer lines.

For the video `Regular8`, 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 visual rhythms only for the unstable video. We present this figure to show the superiority of our strategy, which uses the mean rows and columns in order to have a better separation of horizontal and vertical movement.

It can be seen from Figure 6a that the visual rhythm with only one row can be negatively influenced by the vertical



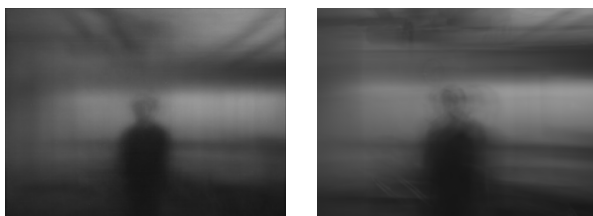
(a) horizontal visual rhythm - mean row (b) horizontal visual rhythm - central row

Figure 6. Visual rhythms for original video `Regular8`.

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.

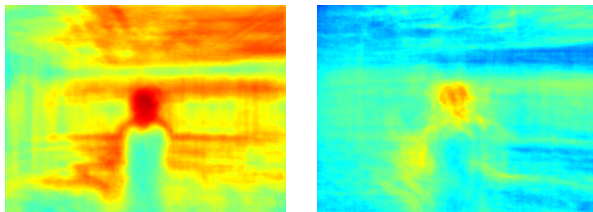
E. Motion Energy Image

In this subsection, we present the results obtained in our experiments with the evaluation based on the motion energy image. The drawback of the average grayscale image, generally used in the literature to compare results visually, becomes clear in the comparison of the results obtained for the video #7. Figures 7 and 8 show the results of the average grayscale image 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.



(a) original video (b) stabilized video

Figure 7. Average grayscale image for video #7.



(a) original video (b) stabilized video

Figure 8. Average image of the colored MEIs for video #7.

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.

IV. CONCLUSIONS, PUBLICATIONS AND DISTINCTIONS

The main objective of this work was to investigate the problem of video stabilization. We then developed and evaluated 2D methods for digital stabilization of videos. The 2D video stabilization process is usually divided into three main steps: estimation of camera motion, removal of unwanted motion, and generation of the corrected video. This work presented five novel methods related to digital video stabilization. Experiments were conducted on three distinct sets of videos.

From the investigation conducted on this work, we have identified some directions that can be explored in future work: (i) construction of a local motion estimation technique for the frames in which the optimization is applied, extending our method to deal with local motion, (ii) development of a new method for removing unwanted motion through a constrained optimization, which obtained the smoothest camera path possible considering a certain minimum amount of frame pixels to be held, and (iii) proposition of objective metrics calculated from the visual representations proposed in our work, using them for the characterization and evaluation of video stabilization.

This work generated the following publications:

- The consensual approach to combining different methods of local features was published in the *Signal, Image and Video Processing* journal (Qualis B1) [39].
- The proposed optimization technique to calculate a new estimate of the corrected motion was published in the *IET Image Processing* journal (Qualis B1) [40].
- The adaptive Gaussian filter was published in the *EURASIP Journal on Image and Video Processing* (Qualis B1) [41].
- The paper related to the motion energy image was published in *The Visual Computer* journal (Qualis A2) [42].
- A paper related to visual rhythms has been submitted to the *EURASIP Journal on Image and Video Processing* (Qualis B1) [43].

This M.Sc. dissertation received the first place in the Competition of Theses and Dissertations of the Institute of Computing of the University of Campinas (UNICAMP). The work also received the first place in XXXII Competition of Theses and Dissertations, organized by the Brazilian Computer Society (SBC), as the best M.Sc. dissertation concluded in 2018.

ACKNOWLEDGMENT

The authors are thankful to CNPq, CAPES and FAPESP for their financial support.

REFERENCES

- [1] Z. Zhao and X. Ma, "Video Stabilization based on Local Trajectories and Robust Mesh Transformation," in *IEEE International Conference on Image Processing*. IEEE, 2016, pp. 4092–4096.
- [2] B.-Y. Chen, K.-Y. Lee, W.-T. Huang, and J.-S. Lin, "Capturing Intention-based Full-Frame Video Stabilization," *Computer Graphics Forum*, vol. 27, no. 7, pp. 1805–1814, 2008.
- [3] H. Guo, S. Liu, S. Zhu, and B. Zeng, "Joint Bundled Camera Paths for Stereoscopic Video Stabilization," in *IEEE International Conference on Image Processing*. IEEE, 2016, pp. 1071–1075.
- [4] S. Liu, B. Xu, C. Deng, S. Zhu, B. Zeng, and M. Gabbouj, "A Hybrid Approach for Near-Range Video Stabilization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.
- [5] B.-H. Chen, A. Kopylov, S.-C. Huang, O. Seredin, R. Karpov, S.-Y. Kuo, K. R. Lai, T.-H. Tan, M. Gochoo, and D. Bayanduuren, "Improved Global Motion Estimation via Motion Vector Clustering for Video Stabilization," *Engineering Applications of Artificial Intelligence*, vol. 54, pp. 39–48, 2016.
- [6] B. Chen, J. Zhao, and Y. Wang, "Research on Evaluation Method of Video Stabilization," in *International Conference on Advanced Material Science and Environmental Engineering*. Atlantis Press, 2016, pp. 253–258.
- [7] J. Kopf, "360 Video Stabilization," *ACM Transactions on Graphics*, vol. 35, no. 6, p. 195, 2016.
- [8] C. Buehler, M. Bosse, and L. McMillan, "Non-Metric Image-based Rendering for Video Stabilization," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2001.
- [9] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum, "Full-Frame Video Stabilization with Motion Inpainting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1150–1163, 2006.
- [10] L. Marcenaro, G. Vernazza, and C. S. Regazzoni, "Image Stabilization Algorithms for Video-Surveillance Applications," in *International Conference on Image Processing*, vol. 1. IEEE, 2001, pp. 349–352.
- [11] A. A. Amanatiadis and I. Andreadis, "Digital Image Stabilization by Independent Component Analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 7, pp. 1755–1763, 2010.
- [12] D. Shukla and R. K. Jha, "A Robust Video Stabilization Technique using Integral Frame Projection Warping," *Signal, Image and Video Processing*, vol. 9, no. 6, pp. 1287–1297, 2015.
- [13] H. Qu, L. Song, and G. Xue, "Shaking Video Synthesis for Video Stabilization Performance Assessment," in *Visual Communications and Image Processing*. IEEE, 2013, pp. 1–6.
- [14] A. Hamza, R. Hafiz, M. M. Khan, Y. Cho, and J. Cha, "Stabilization of Panoramic Videos from Mobile Multi-camera Platforms," *Image and Vision Computing*, vol. 37, pp. 20–30, 2015.
- [15] J. Bai, A. Agarwala, M. Agrawala, and R. Ramamoorthi, "User-Assisted Video Stabilization," in *Computer Graphics Forum*, vol. 33, no. 4. Wiley Online Library, 2014, pp. 61–70.
- [16] M. Okade and P. K. Biswas, "Video Stabilization using Maximally Stable Extremal Region Features," *Multimedia Tools and Applications*, vol. 68, no. 3, pp. 947–968, 2014.
- [17] S. W. Kim, S. Yin, K. Yun, and J. Y. Choi, "Spatio-temporal Weighting in Local Patches for Direct Estimation of Camera Motion in Video Stabilization," *Computer Vision and Image Understanding*, vol. 118, pp. 71–83, 2014.
- [18] S. Liu, L. Yuan, P. Tan, and J. Sun, "Steadyflow: Spatially Smooth Optical Flow for Video Stabilization," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4209–4216.
- [19] M. Grundmann, V. Kwatra, and I. Essa, "Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2011, pp. 225–232.
- [20] J.-Y. Chang, W.-F. Hu, M.-H. Cheng, and B.-S. Chang, "Digital Image Translational and Rotational Motion Stabilization using Optical Flow Technique," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 1, pp. 108–115, 2002.
- [21] S. Ertürk, "Real-Time Digital Image Stabilization using Kalman Filters," *Real-Time Imaging*, vol. 8, no. 4, pp. 317–328, 2002.
- [22] R. Jia, H. Zhang, L. Wang, and J. Li, "Digital Image Stabilization based on Phase Correlation," in *International Conference on Artificial Intelligence and Computational Intelligence*, vol. 3. IEEE, 2009, pp. 485–489.
- [23] S.-J. Ko, S.-H. Lee, and K.-H. Lee, "Digital Image Stabilizing Algorithms based on Bit-Plane Matching," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 617–622, 1998.
- [24] S. Kumar, H. Azartash, M. Biswas, and T. Nguyen, "Real-Time Affine Global Motion Estimation using Phase Correlation and its Application for Digital Image Stabilization," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3406–3418, 2011.
- [25] C.-T. Lin, C.-T. Hong, and C.-T. Yang, "Real-Time Digital Image Stabilization System using Modified Proportional Integrated Controller," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 427–431, 2009.
- [26] C. Morimoto and R. Chellappa, "Fast Electronic Digital Image Stabilization," in *13th International Conference on Pattern Recognition*, vol. 3. IEEE, 1996, pp. 284–288.
- [27] Y. G. Ryu and M. J. Chung, "Robust Online Digital Image Stabilization based on Point-Feature Trajectory without Accumulative Global Motion Estimation," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 223–226, 2012.
- [28] J. Yang, D. Schonfeld, C. Chen, and M. Mohamed, "Online Video Stabilization based on Particle Filters," in *International Conference on Image Processing*. IEEE, 2006, pp. 1545–1548.
- [29] Q. Zheng and M. Yang, "A Video Stabilization Method Based on Inter-frame Image Matching Score," *Global Journal of Computer Science and Technology*, vol. 17, no. 1-F, 2017.
- [30] N. Bhowmik, V. Gouet-Brunet, L. Wei, and G. Bloch, "Adaptive and Optimal Combination of Local Features for Image Retrieval," in *International Conference on Multimedia Modeling*. Springer, 2017, pp. 76–88.
- [31] C. Jia and B. L. Evans, "Online Motion Smoothing for Video Stabilization via Constrained Multiple-model Estimation," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 25, 2017.
- [32] S. Liu, M. Li, S. Zhu, and B. Zeng, "CodingFlow: Enable Video Coding for Video Stabilization," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3291–3302, 2017.
- [33] M. R. Souza, "Digital Video Stabilization: Algorithms and Evaluation," Master's thesis, Institute of Computing, University of Campinas, 2018.
- [34] R. Stephens, *Essential Algorithms: A Practical Approach to Computer Algorithms*. John Wiley & Sons, 2013.
- [35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [36] M. J. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," *The Computer Journal*, vol. 7, no. 2, pp. 155–162, 1964.
- [37] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphics Gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.
- [38] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled Camera Paths for Video Stabilization," *ACM Transactions on Graphics*, vol. 32, no. 4, p. 78, 2013.
- [39] M. R. Souza and H. Pedrini, "Combination of Local Feature Detection Methods for Digital Video Stabilization," *Signal, Image and Video Processing*, vol. 12, no. 8, pp. 1513–1521, 2018.
- [40] M. R. Souza, L. F. R. da Fonseca, and H. Pedrini, "Improvement of Global Motion Estimation in Two-Dimensional Digital Video Stabilisation Methods," *IET Image Processing*, vol. 12, no. 12, pp. 2204–2211, 2018.
- [41] M. R. Souza and H. Pedrini, "Digital Video Stabilization Based on Adaptive Camera Trajectory Smoothing," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 37, 2018.
- [42] —, "Motion Energy Image for Evaluation of Video Stabilization," *The Visual Computer*, pp. 1–13, 2017.
- [43] —, "Visual Rhythms for Qualitative Evaluation of Video Stabilization," *EURASIP Journal on Image and Video Processing*, 2018, (submitted).