A performance evaluation framework for video stabilization methods

Wilko Guilluy
L2TI
University Paris 13
Villetaneuse, France
wilko.guilluy@univ-paris13.fr

Azeddine Beghdadi
L2TI
University Paris 13
Villetaneuse, France
azeddine.beghdadi@univ-paris13.fr

Laurent Oudre
L2TI
University Paris 13
Villetaneuse, France
laurent.oudre@univ-paris13.fr

***Abstract*—This study aims at discussing both objective and subjective aspects of Video Stabilization Quality Assessment (VSQA). A corpus of various degraded videos representing different challenging scenarios and their corresponding processed outputs obtained by four representative methods of Video Stabilization techniques is dedicated to this study. The objective evaluation is restricted to four common VSQA metrics and a new one. The subjective experiments were performed in a laboratory controlled environment using pair wise comparison ranking protocol. Through the obtained results it is shown that the performance evaluation of VS methods is far from being understood and there is still a way to go before satisfactory approaches emerge. This contribution is a first step into the direction of filling this gap by proposing a video stabilization quality assessment methodology.**

I. Introduction

Recent years have seen a significant increase in the production of videos due to the increased availability of affordable cameras, notably camera-equipped smart-phones. Over the last few years, smart-phones have been equipped with more and more powerful sensors that have dramatically improved the quality of the videos in terms of resolution and sensitivity. However, these videos may still suffer from distortions and artifacts, that are not related to the sensors, but rather to the acquisition conditions and the user [1]. Among several considerations, the movements of the camera have important repercussions on the video quality. The acquired video often exhibits instabilities due to uncontrolled camera shakes or intentional cameramen movements that are a source of visual discomfort for viewers and may affect the performance of some video processing and analysis tasks [2].

This annoying distortion could be removed or reduced by using hardware or software solutions for video stabilization (VS) purpose. Some hardware solutions, such as tripods or optical stabilizers in cell-phones[3], have been proposed. However, those solutions are either too cumbersome for everyday use or too limited in the amount of motion that can be compensated. In addition, such methods cannot always be applied in some applications, for instance in the case of video guided surgery. Digital video stabilization techniques are then potential alternative solutions to remove or reduce these spatio-temporal distortion and visual discomfort, either in real-time or at the post-capture processing stage. VS consists in estimating and removing the *unwanted* movements in the video in order to reduce the visual discomfort while preserving a good perceptual quality of the observed video frames.

The definition of the intentional and unintentional parts of the movements observed in the video is inherently subjective, as the formal definition would require a priori knowlegde on the original intention of the cameraman during the video capture, which is often unknown or difficult to predict. This is why the performance evaluation of the digital video stabilization methods remains an unsolved problem and an active research field.

While many video stabilization methods have been proposed a little attention has been paid to video stabilization quality assessment (VSQA). Very often the quality of the processed video is evaluated using some subjective and intuitive criteria or by using some simple quantitative measures. However, these quantitative measures do not exploit any knowledge nor well defined model of the visual discomfort due to video instability. Furthermore, these measures are based on some features or characteristics of motion that are considered as the main origins of such annoying distortion. This study aims at discussing the existing VSQA metrics and propose a framework for developing a methodology for effective VSQA. Here, we restrict the study to some simple quantitative measures and some VS representative methods.

II. Performance evaluation

Over the two last decades many methods for video stabilization have been proposed in the literature. However, some open problems have not yet been thoroughly addressed. Indeed, there are no well-defined criteria for assessing the quality of video stabilization. One of the main reasons is the lack of an established and widely accepted perceptual model to quantify the visual discomfort that may result from camera movement or other electronic instabilities. To the best of our knowledge there has been very few studies dedicated to performance evaluation of video stabilization methods [4]-[5].

The lack of such studies is mainly due to the fact that video instability, like other spatio-temporal distortions and artifacts, is very difficult to model. Indeed, the way this distortion affects the perceived quality is misunderstood and there is no way on how to quantify, in an effective way, the effect of this distortion on the overall quality of the video. In the absence of a well accepted methodology for comparing the existing video stability methods, we perform both subjective evaluation experiment and objective measurements to compare the performance of some representative and available video stabilization methods. In the following we provide some details on the experimental setup, the subjective evaluation protocol, some VSQA metrics and the databases used in this study.

A. Experimental framework and database description

The performance of four representative video stabilization methods is evaluated using the database assembled by Koh et al. [6] from different publications [7], [8], [9], [10], [11], [12] and available videos. This database contains the most challenging problems for video stabilization. This set of videos is organized into 7 categories: “Simple”, “Object”, “Depth”, “Rolling Shutter”, “Crowd”, “Driving” and “Running”, according to the video content and the challenges that it presents for video stabilization methods.

Videos from the “Simple” category contain smooth depth differences and slow camera motion, which do not present any particular challenge to most video stabilization methods. Videos from the “Object” category contain large moving objects in the foreground, which represent a very challenging problem as it is very difficult to discriminate between camera-induced motion and motion resulting from moving objects. Therefore, any common VS method that could not discriminate between these movements may produce inconsistent results. Videos from the “Depth” category contain strong depth discontinuities. This is another very challenging problem for VS methods based on 2D motion models since the depth aspect is not properly taken into account in these approaches and especially in the case of important depth. The “Rolling Shutter” category consists of videos taken using a camera that captures the video row by row. In the case of fast camera motion this may produce noticeable video distortion due this scan-line acquisition system. The videos in this category contain such artifacts, usually caused by fast lateral motion which tilts vertical structures of the scene. Videos from the category “Crowd” contain many independent movements and occlusions, which makes the process of determining the camera motion to stabilize more complicated. The “Driving” sequences are taken from videos embarked on moving vehicles. These videos contain a main steady forward motion disturbed by high frequency shakes of variable intensity, as well as important depth differences and occlusions or moving objects. The ”Running” sequences contain excessive shakes that are difficult to correct while maintaining a good video resolution as well as important depth differences.

B. Methods

In this work, we propose to assess the performances of four standard video stabilization methods. These methods have been chosen for their representativeness of the current approaches that appear in the literature. They are also all free of charge and their source-code or software are available, this allows to conduct a fair comparison of the considered VS methods. In the following, we provide a brief description of the VS techniques used in this study.

- Deshaker [13] is a free plugin that can be used within the VirtualDub software. It is a fast, free and ready-to-use tool that stabilizes horizontal/vertical panning, rotation and zooming. The method assumes that the camera movement between two successive frames can be modelled as a 2D transform (homography, affinity...). Its interface offers a large number of settings and parameters that can be useful for advanced users. In the following experimental setting, we have used the default parameters of these softwares. It uses a block-matching algorithm to determine movements within the video with a coarse-to-fine approach to handle large movements. Default settings uses 30-pixel blocks, 30% of the coarser frame as the initial search range then 4 pixels for the search refinement at each iterations, ending with the frame at half the original scale. Motion vectors are discarded if either the maximum pixel difference or the combined differences between blocks are above given thresholds respectively 20 and 300), or if a pixels move too far in a direction that does not fit with the camera model (with a threshold of 4 pixels). The correction applied is computed by minimizing the squared correction and the motion acceleration, with a maximum correction for both panning, zoom and rotation (set to 15 degrees for panning, 15% for zoom and 5 degrees for the rotation). It also allows to consider rolling shutter effect when determining and correcting the camera motion.
- Youtube Stabilizer [8], [14] is arguably the most popular video stabilization software. Approximately 300 hours of videos are uploaded in Youtube each minute, and the Youtube Stabilizer is a routine discretionary option in the upload process. Similarly to Deshaker, the method assumes a simple 2D geometrical transform between two successive frames, but includes a L1-regularization in the smoothing process, that mimics the motions produced by professional cameramen. The strength of the stabilization is finely tuned so as to ensure that the region or subject-of-interest is always visible in the stabilized video. It also uses homography mixtures to detect and correct rolling shutter artifacts.
- Sanchez et al. method [15] is a recent video stabilization method, publicly available on IPOL [16]. This method assumes a 2D homography transformation between two successive frames involving 8 parameters. It is based on a local smoothing of the transform parameters used in the computation of the stabilized video.
- Koh et al.method [6] is also a recent method which, unlike other VS approaches, does not assume any 2D transformation between consecutive frames. Instead, this approach attempts to provide a plausible and perceptually satisfying correction, that is not based on the geometrical

reality of the scene. This method also includes an explicit rolling shutter removal step, and is able to handle large objects in the foreground.

C. Database and experimental setup

We use pairwise comparisons to determine the preferences between different video stabilization techiques. We selected a dataset comprised of 5 videos from each of the 7 categories in the previously mentioned database in order to perform visual tests, for a total of 35 videos. In order to facilitate the visual tests, shorter videos were chosen and truncated to 10 seconds short when they exceed this duration. The 35 videos from the dataset were treated with the stabilization algorithms described above. The original videos were also included to account for the side effect of the VS methods that may degrade the input videos, resulting in 5 versions of each video. Observers were shown two different versions of the same video, and had to choose which one they preferred, according to their own criteria. They also had the option to indicate no preference between the two. Each observer was shown 50 randomly selected pairs of videos, with each of the 35 videos shown at least once. Each pair consists on two randomly chosen versions the video. Videos could be replayed once, but users were instructed that they could chose their preference-choice at any time. In doing so, the negative effects of visual fatigue and the reduction of visual attention are minimized. The tests were performed on eighteen observers at the Laboratoire de Traitement et Transport de l'Information(L2TI) on a calibrated LCD monitor in a controlled environment as the one used for image quality assessment and described in [17].

D. Objective quality assessment

There are very few objective measures that have been used for video-stabilization quality assessment. Some of the proposed measures rely on the output of one of the VS workflow blocks such as the motion estimation, motion compensation or rendering blocks. The idea is then to use the quantitative outputs, for example the key points matching accuracy, at the level of this block as a VSQA index. However, such approach is somehow biased in the sense that it would ignore the other blocks and gives more weight to only the selected block. The other solution is to construct a ground truth by using a synthetic distorted video using a predefined flow field. The idea is then to apply VS method and compare the output to the original free-distortion video. The flow field estimation error could be then used as an objective VSQA metric a s done in [18] for optical flow performance evaluation. The main difficulty for quantifying the quality of the stabilized video is that it is very difficult to automatically identify and discriminate the various video instabilities.

Here, we focus on five common VSQA metrics, namely : Interframe Transformation Fidelity (ITF), Average Speed (AvSpeed), Average Acceleration (AvAcc), Average Percentage of Conserved Pixels (AvPCP) and Inter-frame Similarity Index (ISI), based on the structural similarity index(SSIM).

The most widely used metric for assessing the performance of video stabilization methods is the Interframe Transformation Fidelity (ITF) index [19], [20]. It is based on the video inter-frame PSNR. Given a video I composed of N_f frames, ITF is expressed as the average inter-frame PSNR.

$$\text{ITF} = \frac{1}{N_f - 1} \sum_{i=1}^{N_f - 1} \text{PSNR}(t), \tag{1}$$

where PSNR(t) is the peak signal-to-noise ratio (in dB) based on the mean-square-error between frames $I(t)$ and $I(t+1)$. The intuitive idea behind this metric is that, if the camera movement is smooth (i.e., stabilized video), the similarity between the consecutive frames should be larger than in the presence of strong camera motion. This metric can also assess the distortions and photometric artefacts that may result from the stabilization process. Note that, if no objects/subjects are in movement in the video and if the stabilization is perfect, the ITF would tend towards infinity.

Another metric that can be used in a similar way is the Structural Similarity Index(SSIM). In order to extend its application to video streams, we define the Interframe Similarity Index (ISI) as the average of the SSIM between successive frames across the video. This new VSQA metric is given by:

$$\text{ISI} = \frac{1}{N_f - 1} \sum_{i=1}^{N_f - 1} \text{SSIM}(t), \tag{2}$$

where SSIM(t) is the structural similarity index between frames $I(t)$ and $I(t+1)$. High values of ISI mean that successive frames are perceptually similar, which is provide better visual comfort for viewer.

Another aspect that could be analyzed when judging the quality of the stabilized video is the local motion. Indeed, the stabilization process acts on the movements present in the video. The stabilization process should produce a smoothing effect of these annoying movements leading to a fluid video sequence. This can be checked by extracting salient feature points in the video and analyzing their displacement along the video. Intuitively, in a properly stabilized video, the movements of these feature points are smooth, i.e. with small speed/acceleration. Considering the i^{th} features point with coordinates $z_i(t)$ in frame $I(t)$, its movement can be characterized by its instantaneous speed $\dot{z}_i(t) = z_i(t+1) - z_i(t)$ or its instantaneous acceleration $\ddot{z}_i(t) = z_i(t+1) - 2z_i(t) + z_i(t-1)$. The Average Speed (AvSpeed) metric is expressed as the average speed of all feature points along the video [21]. If a total of N_p feature points are extracted in the video, the AvSpeed metrics is given by:

$$\text{AvSpeed} = \frac{1}{N_p(N_f - 1)} \sum_{i=1}^{N_p} \sum_{t=1}^{N_f - 1} \|\dot{z}_i(t)\|_2 \,, \tag{3}$$

and is defined as the average quantity of movement of the feature points: it should be as low as possible. In the literature, several feature points have been proposed for motion computation (SURF, SIFT, KLT, sparse optical flow...). In our

		original	Deshaker [13]	Youtube [8]	Sanchez et al. [15]	Koh et al. [6]
crowd	*ITF (dB)*	19.03	20.35	18.24†	**22.67**	22.42
	AvSpeed (px/fr)	3.83	2.71	5.78†	2.81	**2.52**
	AvAcc (px/fr^2)	1.12	0.40	0.90	0.43	**0.33**
	AvPCP (%)	100	**78**†	69†	54†	74†
	ISI	0.62	0.70	0.60†	**0.78**	0.72
depth	*ITF (dB)*	21.67	22.05	22.7	23.8	**24.01**
	AvSpeed (px/fr)	3.50	2.55	2.99	2.65	**2.21**
	AvAcc (px/fr^2)	2.84	0.79	2.20	0.58	**0.53**
	AvPCP (%)	100	83†	**92**†	64†	83†
	ISI	0.70	0.76	0.75	0.80	**0.82**
driving	*ITF (dB)*	23.18	20.88†	22.03†	22.78†	**23.11**†
	AvSpeed (px/fr)	2.56	1.39	1.36	1.42	**1.23**
	AvAcc (px/fr^2)	2.85	0.86	0.94	0.91	**0.53**
	AvPCP (%)	100	**85**†	79†	62†	78†
	ISI	0.69	0.69	0.75	0.75	**0.76**
object	*ITF (dB)*	21.17	22.64	23.21	**24.95**	23.88
	AvSpeed (px/fr)	3.54	1.99	3.07	3.81†	**1.89**
	AvAcc (px/fr^2)	2.21	0.93	0.88	1.38	**0.64**
	AvPCP (%)	100	**80**†	70†	49†	79†
	ISI	0.68	0.76	0.76	**0.85**	0.79
rollingShutter	*ITF (dB)*	20.65	24.09	**29.75**	28.62	26.43
	AvSpeed (px/fr)	7.29	2.09	2.32	**1.46**	1.70
	AvAcc (px/fr^2)	6.16	1.15	0.50	0.90	**0.40**
	AvPCP (%)	100	**69**†	**69**†	60†	63†
	ISI	0.59	0.77	**0.88**	0.86	0.86
running	*ITF (dB)*	18.46	22.9	22.65	**26.57**	25.88
	AvSpeed (px/fr)	7.14	1.91	3.32	1.74	**1.67**
	AvAcc (px/fr^2)	3.14	0.74	0.96	0.78	**0.41**
	AvPCP (%)	100	**61**†	52†	37†	57†
	ISI	0.56	0.69	0.76	**0.85**	0.75
simple	*ITF (dB)*	24.48	27.87	**31.64**	31.33	28.64
	AvSpeed (px/fr)	3.24	1.29	**0.96**	1.06	1.34
	AvAcc (px/fr^2)	2.89	0.61	0.40	0.48	**0.36**
	AvPCP (%)	100	**83**†	72†	72†	**83**†
	ISI	0.74	0.88	**0.93**	0.93	0.88
average	*ITF (dB)*	21.23	22.97	24.32	**25.82**	24.91
	AvSpeed (px/fr)	4.44	1.99	2.83	2.14	**0.1.79**
	AvAcc (px/fr^2)	3.24	1.29	0.97	0.78	**0.46**
	AvPCP (%)	100	**77**†	72†	57†	**74**†
	ISI	0.65	0.75	0.78	**0.83**	0.79

TABLE I

RESULTS ON THE OBJECTIVE METRICS. THE SCORES DENOTED WITH $\dagger$ CORRESPOND TO CASES WHERE THE STABILIZED VIDEO HAS WORSE OBJECTIVE QUALITY LEVEL THAN THAT OF THE ORIGINAL VIDEO.

experiment, we make use of the KLT descriptors thanks to their relatively low computational cost and efficiency.

The analysis of the quantity of movements present in the video may not be sufficient to assess the qualitative aspects of video stabilization. Indeed, the perception of the movement is not only linked to the quantity but also to the type of movement present in the video. In particular, professional cameramen use hardware solutions (such as steadycam), that tend to produce movements with linear displacement or speed [22]. Furthermore, if the cameraman is attempting to follow a subject of interest, the observed motion is intentional and should not be removed. For these reasons, several authors have considered that stabilization should be assessed according to the acceleration rather than the speed of feature points. The Average Acceleration (AvAcc) metric [23] could be expressed as:

$$\text{AvAcc} = \frac{1}{N_p(N_f-2)} \sum_{i=1}^{N_p} \sum_{t=2}^{N_f-1} \left\| \ddot{z}_i(t) \right\|_2 , \quad (4)$$

This metric quantifies the average acceleration of the feature points. For a good stabilized video, this measure is low.

Video stabilization naturally induces a resolution loss in the processed video. Indeed, in case of strong stabilization, the inverse 2D or 3D transform applied on frames often create unknown area in the video that cannot be interpolated without additional information. To circumvent this limitation, one has to apply some post-processing solutions such as cropping, zooming or video resizing, so as to remove those blank areas. The loss of resolution, if large, may produce an annoying effect and can therefore be considered as a criterion of evaluation [15] . One way to quantify this effect is to express the ratio of pixels that survive the VS process. Given the original video I

	M1	M2	M3	M4	M5
M1	-	5.5	7	9	4.5
M2	12.5	-	11	10	7.5
M3	11	7	-	8.5	7.5
M4	9	8	9.5	-	6
M5	13.5	10.5	10.5	12	-

TABLE II

SAMPLE PREFERENCE MATRIX AGGREGATED ON ALL VIDEOS FOR 18 OBSERVERS. SCORE IN CASE (i, j) CORRESPONDS TO THE AVERAGE NUMBER OF OBSERVERS THAT PREFERRED METHOD i OVER METHOD j. M1 = NO STABILIZATION, M2 = DESHAKER [13], M3 = YOUTUBE [8], M4 = SANCHEZ ET AL. [15], M5 = KOH ET AL. [6]

.

and a stabilized video $\tilde{I}$, the Average Percentage of Conserved Pixels (AvPCP) could exeressed as the following ratio:

$$\text{AvPCP} = \frac{100}{N_f} \sum_{t=1}^{N_f} \frac{\text{res}(\tilde{I}_t)}{\text{res}(I_t)}, \quad (5)$$

where res(.) is the resolution (in pixels) (i.e.res$(I_t) = MN$ if I_t if of size $N \times M$). Obviously, a well stabilized video should have high AvPCP value so as to minimize the fraction of lost pixels during the stabilization process. It is worth noticing that the interpretation of this metric greatly depends on the level of stabilization: in particular, note that this metric equals 100% if no stabilization is performed. To conclude, this metric should only be used in addition to other performance metrics or to evaluate methods with the same level of stabilization.

III. RESULTS AND DISCUSSION

The performance evaluation of the 4 VS methods has been done on a set of 7 video categories representing various challenging scenarios as mentioned previously. In the following we provide a brie discussion on some preliminary results by considering both subjective and objective aspects.

A. *Objective performance evaluation*

Table I displays the values of the objective metrics obtained by the methods and video categories described in Section II. The first observation is that the method by Koh et al. obtains the best general performances according to the AvSpeed and AvAcc criteria, and second best on the ITF and ISI metrics. It outperforms other methods on all categories for at least one metric of evaluation. Whereas, Deshaker obtains the lowest performances for both ITF and ISI, but the best result according to the AvPCP metric. Using the default settings, this method specifically avoids excessive corrections in order to maintain video resolution. As expected and described in Section II-D, the AvPCP is in contradiction with other metrics: there is a compromise to make between keeping the video resolution high and perform a severe stabilization. Interestingly, the metric values are sometimes better on the original unstabilized video than on the stabilized one. This phenomenon is especially observed with Youtube and Sanchez et al. methods, for instance on the Driving or Crowd categories. In fact, since these methods are based on 2D geometrical models, they sometimes fail at dealing with videos in which the camera movement lies in a 3D space or includes parallax effects. This causes erratic stabilizations that mis-estimate the camera movement and actually create additional distortions in the video, which are hard to estimate as those metrics show good results on image similarity metrics. Apart from the AvPCP metrics, it appears that all metrics are coherent with each other: although their ranking is somewhat different. Globally, it could be noticed that the metric values are of the same order for the considered dataset.

B. *Subjective performance evaluation*

Table II presents the sample preference matrix of the subjective pair comparison tests, averaged on all videos. The method by Koh et al. obtains the best overall results : it is preferred to any other methods in two-thirds of cases. This result is coherent with the results obtained with the IFT, AvSpeed and AvAcc metrics. The second best method is Deshaker, which may appear surprising according to the results of Table I. In fact, the fact that Deshaker does not perform strong stabilizations and keeps a reasonable video resolution has for consequence that this methods never produce aberrant results, on the contrary to Youtube or Sanchez et al. Thus, strong stabilization is not the only criterion seeked by the observer, that prefers a video with acceptable camera movements than a fully stabilized video with distortions. The two last methods are Youtube and Sanchez et al. which perform a good stabilization (see Table I) with high image similarity metrics but fail on several videos and tend to produce over-cropping. Interestingly, no stabilization is sometimes preferred to any stabilization, which confirms that video stabilization is a processing step that can in fact cause distortions. It should be noted that both Youtube and Sanchez et al. seem to perform well in the "simple" category, and that as this database is representative of the range of challenges for video stabilization, it may not be representative of the usual type of scenes for which these methods are optimized.

IV. CONCLUSION

Through this study it has been shown that the performance evaluation of video stabilization methods is far from being an easy subject. The few objective quality assessment measures that have been proposed in the literature do not include any knowledge of the effect of video instabilities on the human visual system. This study was not intended to propose such models but to encourage people to work in this direction and propose models to account for visual discomfort that may result from video instabilities. To the best of our knowledge there have been no complete subjective and objective studies dedicated to VSQA. This contribution is a first step into the direction of filling this gap by proposing a video stabilization quality assessment methodology.

REFERENCES

[1] M. L. Gleicher and F. Liu, "Re-cinematography: Improving the camera dynamics of casual video," in *Proceedings of the ACM International Conference on Multimedia*, 2007, pp. 27–36.

[2] S. Megrhi, M. Jmal, W. Souidene, and A. Beghdadi, "Spatio-temporal action localization and detection for human action recognition in big dataset," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 375–390, 2016.

[3] J.-M. Koo, M.-W. Kim, and B.-K. Kang, "Optical image stabilizer for camera lens assembly," Patent, 2009, uS Patent 7,489,340.

[4] M. Niskanen, O. Silvén, and M. Tico, "Video stabilization performance assessment," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2006, pp. 405–408.

[5] C. Morimoto and R. Chellappa, "Evaluation of image stabilization algorithms," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1998, pp. 2789–2792.

[6] Y. J. Koh, C. Lee, and C.-S. Kim, "Video stabilization based on feature trajectory augmentation and selection and robust mesh grid warping," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5260–5273, 2015.

[7] A. Goldstein and R. Fattal, "Video stabilization using epipolar geometry," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 5, p. 126, 2012.

[8] M. Grundmann, V. Kwatra, and I. Essa, "Auto-directed video stabilization with robust l1 optimal camera paths," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 225–232.

[9] F. Liu, M. Gleicher, J. Wang, H. Jin, and A. Agarwala, "Subspace video stabilization," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 1, p. 4, 2011.

[10] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled camera paths for video stabilization," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 78, 2013.

[11] ——, "Steadyflow: Spatially smooth optical flow for video stabilization," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 4209–4216.

[12] A. Karpenko and D. Jacobs, "Ctsr 2011-03 digital video stabilization and rolling shutter correction using gyroscopes," 2011.

[13] G. Thalin, "Deshaker," *http://www.guthspot.se/video/deshaker.htm*, 2008.

[14] M. Grundmann, V. Kwatra, D. Castro, and I. Essa, "Effective calibration free rolling shutter removal," *IEEE ICCP*, 2012.

[15] J. Sánchez and J.-M. Morel, "Motion smoothing strategies for 2D video stabilization," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 219–251, 2018.

[16] J. Sánchez, "Comparison of Motion Smoothing Strategies for Video Stabilization using Parametric Models," *Image Processing On Line*, vol. 7, pp. 309–346, 2017.

[17] M. A. Qureshi, A. Beghdadi, and M. A. Deriche, "Towards the design of a consistent image contrast enhancement evaluation measure," *Sig. Proc.: Image Comm.*, vol. 58, pp. 212–227, 2017. [Online]. Available: https://doi.org/10.1016/j.image.2017.08.004

[18] A. Beghdadi, M. Mesbah, and J. Monteil, "A fast incremental approach for accurate measurement of the displacement field," *Image Vision Comput.*, vol. 21, no. 4, pp. 383–399, 2003. [Online]. Available: https://doi.org/10.1016/S0262-8856(03)00014-3

[19] L. Marcenaro, G. Vernazza, and C. S. Regazzoni, "Image stabilization algorithms for video-surveillance applications," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2001, pp. 349–352.

[20] J. Xu, H.-w. Chang, S. Yang, and M. Wang, "Fast feature-based video stabilization without accumulative global motion estimation," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, 2012.

[21] Z. Wang and H. Huang, "Pixel-wise video stabilization," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 15 939–15 954, 2016.

[22] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3D video stabilization," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, p. 44, 2009.

[23] C. Tang and R. Wang, "Local subspace video stabilization," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.