Video stabilization: challenges, methods and evaluation

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

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

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

***Abstract*— In this paper we discuss some of the challenges that video stabilization methods face. We present a new method to select the most relevant feature trajectories to use in video stabilization algorithms and discuss the different methods used to evaluate video stabilization methods.**

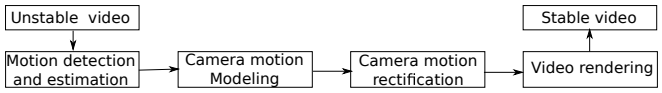

Fig. 1. Flowchart of video stabilization process

I. INTRODUCTION

Recent years have seen a marked increase in the production of amateur videos. While more powerful sensors have improved the quality of the videos, there are still noticeable differences between the perceptual quality of amateur and professional video captures[1], particularly regarding the quality of camera motion. Many videos often exhibits instabilities due to unintentional camera movements that are a source of visual discomfort for viewers and may affect the performance of some video analysis tasks. Video stabilization operates in several interdependent steps, as shown in figure 1. First, the video motion field is estimated. Secondly, the original camera path is computed. The camera path is then rectified to obtain more coherent movements and a new stable video is rendered. While each step has its own set of difficulties, the most challenging aspects of video stabilization lie in the estimation of the camera motion and the evaluation of the stabilization. Firstly, all detected movements do not give good information about the camera motion. Movements caused by moving objects are not the result of the camera motion and can lead to errors if not removed. Secondly, the camera moves in 3D, but videos do not usually contain depth information, making 3D models difficult to compute. 2D models are more robust and work faster, but can lead to distortions in the presence of large depth differences. Other non-geometrical models have been proposed to strike a balance between complexity and reliability[2]. Choosing the right model is therefore a challenging task. Finally, video stabilization evaluation is a multi-criterion problem: visual discomfort due to the camera motion, artifacts caused by the stabilization process such as resolution loss and distortions all contribute to the quality of the output video. Additionally, visual discomfort is inherently subjective, making objective evaluation difficult. We propose to address the selection of movements useful to the camera motion evaluation through a feature trajectory selection method. We then review different quantitative stabilization methods and compare them to a qualitative evaluation by pair-wise comparison. We finish with some concluding remarks.

II. FEATURE TRAJECTORY SELECTION

The most commonly used method to identify and remove movements caused by objects rather than the camera motion are based on the RANSAC algorithm. However, this only considers one pair of frames at a time, and needs the motion to fit a determined motion model. Feature point tracking offer the advantage of enabling a temporal analysis of the different features identified. Our method uses the KLT tracker, although any feature tracking scheme could be used. Our goal is to exploit the evolution of the trajectories over time, and avoid the reliance on a given motion model. We differentiate the object and camera motion using the duration of the trajectory and the characteristics of its movements.

The feature trajectories selection strategy proposed in this article is based on the following steps:

1) We analyze each feature trajectory on a local time-window, in order to account for its duration and movement properties. Two local weights are defined that rank each trajectory according to its duration and its adequacy with the movements observed on a time-window centered on a given frame.
2) We combine all local weights in order to form a global trajectory weight that accounts for the phenomenon observed during the whole duration of the trajectory.
3) We select the feature trajectories with the largest weights to estimate the camera motion parameters.

A. Duration characteristics of features trajectories

Feature trajectories that span too few frames are likely to be unreliable. In most cases, they correspond to feature points that are not salient enough to be tracked, or to moving objects that do not remain in the scene for a long time.

To this end we consider a sliding time window and compute a duration weight that accounts for the local duration of the considered trajectory. This weight corresponds to the percentage of frames within the temporal window of interest for which the feature is detected. It is a local score that provides a temporal assessment of the reliability of the feature trajectory.

B. Motion characteristics of features trajectories

A second criterion related to the motion characteristics of the features trajectories is introduced to discriminate between static and moving objects in the distorted video. However, without knowledge of the scene characteristics, deriving the most appropriate motion model for the video is an ill-posed problem. Therefore, instead of using RANSAC, which is based on geometric models, we propose to identify the dominant movement in the video without assuming a given motion model, by using a projection in a low-rank subspace[3]. We concatenate the frame-to-frame motion of the different features into a single matrix, which is analyzed locally using the SVD decomposition. This allows us to plot the contributions of the different trajectories to the dominant movement within the considered time-lapse. Intuitively, all feature trajectories belonging to static objects or background, should have similar contributions to this dominant movement. On the contrary, moving objects or pixels should have different contributions and thus be identifiable. Thus the mode of the contribution is considered the best motion, and trajectories are weighted according to the distance between their contribution and the mode.

C. Combination process

Both weights are multiplied together for each frame and trajectories, and the mean weight of each trajectory is then computed to obtain a global weight. Trajectories are then rejected or retained by thresholding the global weights. Some results for both the elimination of feature points and the comparison with a RANSAC-based method and the Youtube Stabilizer can be found on our webpage[1].

III. VS Evaluation

Over the two last decades many methods for video stabilization have been proposed. However, there are no agreed-upon criteria for assessing the quality of video stabilization. One of the main reasons is that video stabilization evaluation is a multi-criteria process. These criteria include the resolution loss, the quality of the camera motion and the perceived distortions. We review several objective metrics to measure these different criteria and compare them to a subjective evaluation using a pair-wide comparison setup.

A. Objective quality assessment

Several metrics have been proposed to measure different aspects of the video stabilization process[4], but so far there is no agreed upon reference for the evaluation of video stabilization. One criterion is the loss of resolution caused by video stabilization. In case of strong stabilization, the stabilization applied often creates unknown area in the video that cannot be interpolated without additional information. The percentage of the undefined areas was used to validate the feature trajectory selection method, resulting in an average gain of 3.67% of undefined areas compare to 5.15% using the standard RANSAC-based method over a dataset of 15 videos. However, since different methods employ different strategies to crop out those areas, we considered the ratio of the resolution before and after stabilization, for our comparison with subjective evaluations. Additionally, The stabilization process should produce a smoothing effect of these annoying movements leading to a fluid video sequence. The impact of the camera motion can be perceived in the motions within the video. Therefore, we track feature points in the original and stabilized sequence to compare the differences between the velocity and acceleration in the original and stabilized video, with the expectation that the stabilization process should produce smoother motion. Finally, metrics such as the ITF (which corresponds to the average of the PSNR) or SSIM measure the similarity between adjacent frames, which are expected to be higher if the video motion are smooth and few distortions occur.

[1] http://www-l2ti.univ-paris13.fr/∼guilly/

B. subjective quality assessment

Subjective quality assessment is often used to validate the results of stabilization. We set up a pair-wise comparison protocol where subjects would select their preference between randomly selected versions of video sequences. The different versions of a sequence include the original sequence and the results of different stabilization methods.

C. comparison

The comparison between subjective and objective evaluation confirm that none of the objective criterion can explain the subjective preferences by itself. Overall, it seems that smooth motion (with low acceleration) is preferable, but that past a certain point resolution is deemed preferable to a more aggressive stabilization. The ITF seems to correlate poorly the the user preferences, and certain on challenging video sequences the user sometimes prefer the unstabilized version of the video.

IV. Concluding Remarks

We have presented a trajectory selection method that improves upon the current methods and is compatible with any motion model assumed by the rest of the stabilization process. We also investigated the different ways to evaluate the stabilization process, as a first step into the direction of filling this gap by proposing a video stabilization quality assessment methodology.

References

[1] 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)*, pp. 225–232, 2011.

[2] K. Karageorgos, A. Dimou, A. Axenopoulos, P. Daras, and F. Alvarez, "Semantic filtering for video stabilization," in *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2017.

[3] F. Liu, M. Gleicher, J. Wang, H. Jin, and A. Agarwala, "Subspace video stabilization," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 1, pp. 4:1–4:10, 2011.

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