Motion Intention Optimization for Multirotor Robust Video Stabilization

Wilbert G. Aguilar
CICTE Research Center
Universidad de las Fuerzas Armadas ESPE
Sangolquí, Ecuador
wgaguilar@espe.edu.ec

Cecilio Angulo
GREC Research Group
Universitat Politècnica de Catalunya
Barcelona, Spain
cecilio.angulo@upc.edu

Jorge A. Pardo
UGT. Universidad de las Fuerzas Armadas ESPE
Sangolquí, Ecuador
japardo@espe.edu.ec

***Abstract*— In this paper we present an optimization algorithm for simultaneously detecting video freeze and obtaining the minimum number of the frame required in motion intention estimation for real time robust video stabilization on multirotor unmanned aerial vehicles. A combination of a filter and a threshold is used to the video freeze detection, and for optimizing the algorithm, we find the minimum number of frames for motion intention estimation without decrease the performance.**

Keywords—Video Stabilization, Motion Smoothing, Warping, RANSAC, Geometric Transform, feature points, Filter.

I. INTRODUCTION

In [1] [2] [3], we present video stabilization algorithms based on Low-pass and Kalman Filters that can compensate the effects of the undesired movements in real time for micro aerial vehicles. However, we have not considered the problem of video freeze. In this paper we present a proposal for video freeze detection and optimization of the motion intention process.

In the approaches of video stabilization, three phases can be distinguished: Motion estimation, Robust cumulated motion estimation, Motion compensation. Motion estimation is the process for determining parameters that relate the frame uncompensated with frame defined as the reference. Previous works on this problem propose two main approaches: one based on the optical flow [4] and the other based on the geometric transformation model [5] [6] [7]. In this article, we use the second proposal. Independently from the approach, feature points detection and description are required, and there are several algorithms to perform these tasks [8] [9]. SIFT [10] (Scale Invariant Feature Transform) and SURF (Speed Up Robust Feature) [11] are the most widely used algorithms for detection, description, and almost computer vision problems [12].

In the robust cumulated motion estimation, a search of correspondences between feature points is carried out as a part of the motion estimation process. The estimated motion parameters is directly dependent on the reliability of computed matched points. RANSAC is a technique commonly used in the literature to estimate mathematical model parameters from a set of points with false correspondences [13] [14] [15]. Since the complete motion sequence must be coherent, it is important to validate the estimated parameters for the global movement and not just for the relative movement between consecutive frames.

Finally, in the motion compensation process, the current frame is warped using parameters obtained by a robust estimation, and generating a stable video sequence.

II. INTER-FRAME MOTION ESTIMATION

Inter-frame motion estimation is a fundamental step in the video stabilization process, where local motion parameters between consecutives frames are calculated. This estimation process determines the mathematical model as a relation between the current frame and the reference frame.

In spite of the several techniques for detecting and matching interest, results presented in [12] show that the computational cost for SURF is considerably lower than SIFT without a robustness reduction in the algorithm. Using the Hessian matrix and the space-scale function [16], SURF algorithm locates waypoints, and the characteristics are described using a 64-dimensional vector. Once vectors descriptors are computed, feature point matching process selects pairs of points with the minimum vector difference between their 64-dimensional descriptors.

The variations between two frames can be expressed mathematically by the geometric transformation that relates feature points from a frame with their correspondences in the second frame [17] [18] [19]. This geometric transformation has

978-1-5386-3123-2/17/$31.00 ©2017 IEEE

a parametric motion model, and is different depending on the used transformation. Common models are: Translation, Affine, Projective Model.

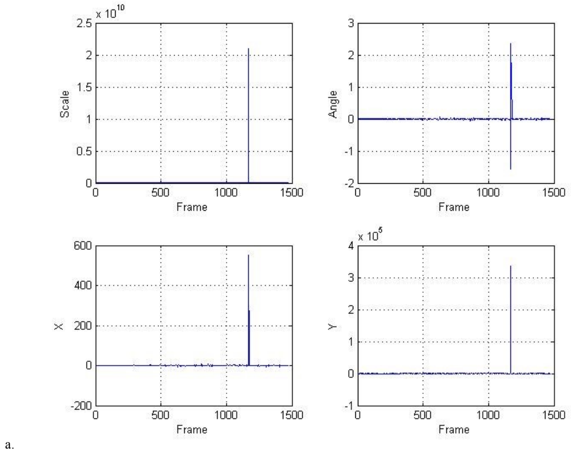
Translation model is the simplest model, referred to the image movement when the capture device motion is only translation in a plane parallel to the image plane (pinhole camera model [20]). In the affine model, there are four parameters to be estimated: two displacements in the plane parallel to the image, as we described in translation model, roll rotation, and scale that is proportional to the motion in roll axis orientation. The projective model is the full motion model, with the mathematical expression for the three possible rotations and translations.

In case of either hand-held digital camera, or monocular vision devices onboard complex dynamic robots, undesired movement and parasitic vibrations in the image are considered significant only about the roll axis. Therefore, the affine model is selected for motion parameters estimation.

There are two additional advantages for this model. First one is referred to a lower computation time. This is because the model depends only on four parameters, compared to the projective model with eight independent parameters. A second advantage is the capability for direct extraction of relevant motion parameters: scale, rotation, roll, and translations in the xy plane.

III. Video freeze detection

Several smoothing methods based on filters have been used in video stabilization algorithms, such as Kalman filtering [22], Gaussian filtering [23], and particle filtering [24]. Once we extract affine transformation parameters (scale, rotation, and translation XY), a low-pass filter is used to get the motion intention.



a.

Fig. 1. Video Freeze Detection

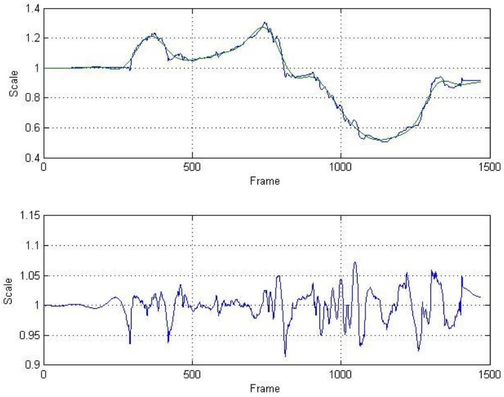
The data collection for the experimental tests has been done with a flying robot. However, the algorithm was executed on a laptop, whose communication with the flying robot was through Wi-Fi. Therefore, video freezing was possible due to connection problems. There is possible to detect and correct freezing problems due to a bad connection, and it is important to eliminate these parameters data from frozen frames before the filter implementation.

When the communication with the flying robot is lost, the estimated parameters increases considerably (Figure 1). After several test, we have set thresholds for to reject the values from frozen frames, which are out of the follow conditions:

- $s > 1.06$
- $\theta > 0.08 rad$
- $t_x > 0.1 * tmax_x$
- $t_y > 0.1 * tmax_y$

where $tmax_x$ and $tmax_y$ are the maximum number of pixel for x and y.

Once we have removed the values from frozen screens, we estimate the motion intention using the low-pass filter to obtain a signal with the undesired high frequency movements. We use this signal in image warping to compensate the vibrations, and simultaneously, keep the intentional motions. In figure 2, it can be appreciated the motion intention signal estimated with the low-pass filter (top figure), and the high frequency signal to be compensated (down figure). The results is comparable with [1] [2] [3].



b.

Fig. 2. Scale intention estimation

IV. Video stabilization optimization

It is important to define the evaluation metric for the optimization of video stabilization.

A. Evaluation Metrics

There are different evaluation metrics for determining the algorithm performance. In the literature, we can find subjective evaluation metrics such as the mean opinion square (MOS), common in the quality evaluation of the compressed multimedia [25]. Other option is to find objective evaluation metrics such as bounding boxes, referencing lines, or synthetic sequences [26]. The advantage of all three objective metrics is that the estimated motion parameters can be directly compared with the ground-truth global motion. However, a widely used method to measure the motion smoothness is the inter-frame transformation fidelity (ITF) [27], whose mathematics expression is:

$$ITF = \frac{1}{N_{frame}-1}\sum_{k=1}^{N_{frame}-1} PSNR(k) \qquad (5)$$

where N_{frame} is the number of video frames and $PSNR(k)$ is the peak signal-to-noise ratio between two consecutive frames $(k, k+1)$ which can be defined as:

$$PSNR(k) = 10 \log_{10} \frac{I_{MAX}}{MSE(k)} \tag{6}$$

$$MSE(k) = \frac{1}{M*N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \| H_{curr} * I_{curr}(i,j) - I_{last}(i,j) \|^2 \tag{7}$$

where I_{MAX} is the maximum pixel intensity and $MSE(k)$ is the mean square error between monochromatic image with size $M * N$.

A second performance objective evaluation metric is the difference between estimated global motion and ground-truth global motion (RMSE):

$$RMSE = \frac{1}{2F} \left(\sqrt{\sum_{j=0}^{F} \left(E_{x,j} - T_{x,j}\right)^2} + \sqrt{\sum_{j=0}^{F} \left(E_{y,i} - T_{y,i}\right)^2} \right) \tag{8}$$

where $E_{x,j}$ and $E_{y,i}$ are the estimated global motion of jth frame in x-axis and y-axis, respectively. $T_{x,j}$ and $T_{y,i}$ are the ground-truth global motion of the jth frame in the x-axis and y-axis, respectively. F denotes the number of frames in a sequence.

Actually, the evaluation criterion is the ratio of translation jitter attenuation defined as:

$$J_c = \frac{1}{F} \sum_{j=1}^{F} {e_j}^2 \tag{9}$$

$$Jitter\ attenuation = \frac{J_{c\_stabilized}}{J_{c\_original}} \tag{10}$$

where J_c represent jitters of a sequence, F is the number of frames in a sequence, and e_j denotes the difference between the derived and optimal translation parameters of the jth frame. Moreover $J_{c\_stabilized}$ and $J_{c\_original}$ is referred to the jitter of a sequence after and before stabilization respectively.

B. Optimization

In order to get a high performance for video stabilization method using a minimum number of frames from a video sequence, we have performed an exhaustive searching by an algorithm that iteratively increase the number of frames used for the motion intention estimation, whose results are plotted in figure 3.

Based on the objective evaluation metric ITF, we have gotten a comparable performance estimating the motion intention with only eleven previous and eleven posterior frames than using the complete sequence.

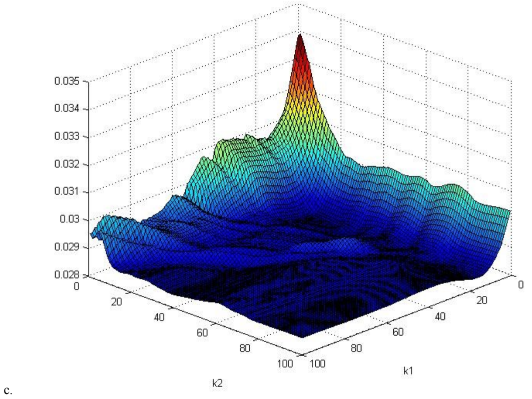


Fig. 3. Minimization of inter-frame transformation fidelity ITF

We modify the criterion for RMSE comparing the N-sample based estimated global motion, with the estimation of the global motion based on all the samples.

$$RMSE = \frac{1}{2F} \left(\sqrt{\sum_{j=0}^{F} \left(N_{x,j} - A_{x,j}\right)^2} + \sqrt{\sum_{j=0}^{F} \left(N_{y,i} - A_{y,i}\right)^2} \right) \tag{11}$$

where the estimated global motion of jth frame in x-axis and y-axis, are respectively represented by: $N_{x,j}$ and $N_{y,i}$ when we use N-samples, and $A_{x,j}$ and $A_{y,i}$ for All-samples.

Using the modified RMSE as a second objective evaluation metric, we have gotten similar results (The four parameters of global motion are separated in Figure 4).

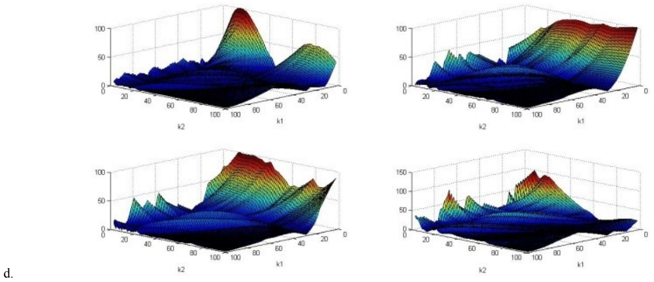


Fig. 4. Minimization of four affine parameters performance

Using the dataset from [1], we can obtain a visual perception of the results obtained for each experimental environment. Since motion estimation is based on feature points, video stabilization is performed around the regions with an agglomeration of these points. Consequently, scene regions remain virtually stationary despite the unstable dynamic motion of capture device.

V. CONCLUSIONS AND FUTURE WORKS

The optimized approach based on motion intention estimation, with video freeze detection, presented in this paper is robust to: Presence of nearby objects, Scenes with moving objects, Scenes frame by frame, Significant displacements, Low frequency videos, High speed displacement, Video freeze.

The low-pass filter for motion estimation is robust into several complex scenes. Matching feature points based on RANSAC is not only referenced to moving objects but to the whole image.

For scene frame to frame, low frequency videos or high speed displacement, the change between two consecutives frames could be considerable, generating a critical problem in video stabilization results. When feature points disappear on the new scene, the algorithm has not reference points or creates false matches.

In the global motion estimation, the algorithm eliminates the data over a threshold generated by video freeze before the filter implementation, becoming robust to considerable changes in the images.

As a future work, we extrapolate video freeze detection to onboard applications as people detection [28] [29], navigation [30] [31] [32], obstacle avoidance [33] [34] [35], and mapping [36].

VI. Acknowledgement

This work is part of the project titled "Perception and localization system for autonomous navigation of rotor micro aerial vehicle in gps-denied environments, VisualNavDrone", 2016-PIC-024, from CICTE Research Center in the Universidad de las Fuerzas Armadas ESPE, directed by Dr. Wilbert G. Aguilar.

References

[1] W. G. Aguilar, and C. Angulo, "Real-Time Model-Based Video Stabilization for Microaerial Vehicles," Neural Process Letters, vol. 43, pp. 459–477, 2016.

[2] W. G. Aguilar, and C. Angulo, "Real-time video stabilization without phantom movements for micro aerial vehicles" EURASIP Journal on Image and Video Processing, vol. 46, 2014.

[3] W. G. Aguilar, and C. Angulo, "Robust video stabilization based on motion intention for low-cost micro aerial vehicles," In Proceedings of 11th International Multi-Conference on Systems, Signals & Devices, pp. 1–6, 2014.

[4] H. C. Chang, S. H. Lai, and K. R. Lu, "A robust and efficient video stabilization algorithm," In Proceedings of IEEE International Conference on Multimedia and Expo, pp. 1:29–32, 2004.

[5] K. Y. Lee, Y. Y. Chuang, B. Y. Chen, and M. Ouhyoung, "Video Stabilization using Robust Feature Trajectories." National Taiwan University, 2009.

[6] J. Yang, D. Schonfeld, and M. Mohamed, "Robust video stabilization based on particle filter tracking of projected camera motion," IEEE Trans. Circuits Syst. Video Technol., vol. 19, pp. 7: 945–954, 2009.

[7] C. Wang, J. H. Kim, K. Y. Byun, J. Ni, and S. J. Ko, "Robust digital image stabilization using the Kalman filter," IEEE Transactions on Consumer Electronics, vol. 55, pp. 1:6-14, 2009.

[8] C. Harris, and M. Stephens, "A Combined Corner and Edge Detector," Proceedings of the 4th Alvey Vision Conference, pp. 147-151, 1988.

[9] J. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 8:679-714, 1986.

[10] D. Lowe, "Object Recognition from Local Scale-Invariant Features," International Conference of Computer Vision, 1999.

[11] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded Up Robust Features," European Conference on Computer Vision, pp. 1:404-417, 2006.

[12] J. Luo, and G. Oubong, "A comparison of SIFT, PCA-SIFT and SURF", International Journal of Image Processing, pp. 143-152, 2009.

[13] M. Fischler, and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, vol. 24, pp. 6:381–395, 1981.

[14] B. Tordoff, and D. W. Murray, "Guided sampling and consensus for motion estimation," European Conference on Computer Vision, 2002.

[15] K. G. Derpanis, "Overview of the ransac algorithm", Technical report, Computer Science, York University, 2010. http://www.cse.yorku.ca/~kosta/CompVis_Notes/ransac.pdf

[16] K. Mikolajczyk, and C. Schmid. "Scale and Anfine Invariant Interest Point Detectors," International Journal of Computer Vision, vol. 60, pp. 1:63-86, 2004.

[17] O. Faugeras, Q. Luong, and T. Papadopoulo, The Geometry of Multiple Images, MIT Press, 2001.

[18] R. Hartley, and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

[19] D. Forsyth, and J. Ponce, Computer Vision, a Modern Approach, Prentice Hall, 2003.

[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, pp. 3:993- 999,2012.

[21] S. J. Kang T. S. Wang, D. H. Kim, A. Morales, and S. J. Ko, "Video stabilization based on motion segmentation," IEEE International Conference on Consumer Electronics ICCE, pp. 416-417, 2012.

[22] C. Wang, J.-H Kim, K.-Y Byun, J. Ni,and S.-J Ko, "Robust digital image stabilization using the kalman filter," IEEE Trans. Consumer Electron., vol. 55, no. 1, pp. 6-14, Feb. 2009.

[23] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H. Shum, "Full-frame video stabilization with motion inpainting," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 28, no. 7, pp. 1150-1163, Jul. 2006.

[24] J. Yang, D. Schonfeld, and M. Mohamed, "Robust video stabilization based on particle filter tracking of projected camera motion," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 7, pp. 945-954, Jul. 2009.

[25] Methods for Subjective Determination of Transmission Quality, ITU-T Recommendation, International Telecommuincation Union. Available: http://www.itu.int/rec/T-REC-P.800-199608-I/en.

[26] Video stabilization using maximally stable extremal region features.

[27] Design and Implementation of Efficient Video Stabilization Engine Using Maximum a Posteriori Estimation and Motion Energy Smoothing Approach.

[28] W. G. Aguilar, M. A. Luna, J. F. Moya, et al "Pedestrian Detection for UAVs Using Cascade Classifiers and Saliency Maps," Lecture Notes in Computer Science, pp. 563–574, 2017.

[29] W. G. Aguilar, M. A. Luna, J. F. Moya, et al "Pedestrian Detection for UAVs Using Cascade Classifiers with Meanshift," IEEE 11th Int. Conf. Semant. Comput, pp. 509–514, 2017.

[30] P. Cabras, J. Rosell, A. Pérez, et al "Haptic-based navigation for the virtual bronchoscopy," IFAC Proc, vol 18, pp. 9638–9643, 2011.

[31] W. G. Aguilar, S. Morales "3D Environment Mapping Using the Kinect V2 and Path Planning Based on RRT Algorithms," Electronics, vol. 5, pp.70, 2016.

[32] W. G. Aguilar, S. Morales, H. Ruiz, V. Abad "RRT* GL Based Optimal Path Planning for Real-Time Navigation of UAVs," Lecture Notes in Computer Science, pp. 585–595, 2017.

[33] W. G. Aguilar, V. P. Casaliglla, J. L. Pólit "Obstacle avoidance based-visual navigation for micro aerial vehicles," Electronics, 2017.

[34] W. G. Aguilar, V. P. Casaliglla, J. L. Pólit, et al "Obstacle Avoidance for Flight Safety on Unmanned Aerial Vehicles," Lecture Notes in Computer Science, pp. 575–584, 2017.

[35] W. G. Aguilar, V. P. Casaliglla, J. L. Pólit "Obstacle Avoidance for Low-Cost UAVs," Proc - IEEE 11th Int Conf Semant Comput ICSC, 2017.

[36] W. G. Aguilar, G. A. Rodríguez, L. Álvarez, et al " Visual SLAM with a RGB-D Camera on a Quadrotor UAV Using on-Board Processing," Lecture Notes in Computer Science, pp. 596–606, 2017.