

RESEARCH

Open Access

# Real-time video stabilization without phantom movements for micro aerial vehicles

Wilbert G Aguilar\* and Cecilio Angulo

## Abstract

In recent times, micro aerial vehicles (MAVs) are becoming popular for several applications as rescue, surveillance, mapping, etc. Undesired motion between consecutive frames is a problem in a video recorded by MAVs. There are different approaches, applied in video post-processing, to solve this issue. However, there are only few algorithms able to be applied in real time. An additional and critical problem is the presence of false movements in the stabilized video. In this paper, we present a new approach of video stabilization which can be used in real time without generating false movements. Our proposal uses a combination of a low-pass filter and control action information to estimate the motion intention.

**Keywords:** Video stabilization; Micro aerial vehicles; Real time; Filter; Motion intention

## Introduction

The growing interest in developing unmanned aircraft vehicles (UAVs) is due to their versatility in several applications such as rescue, transport, or surveillance. A particular type of UAV that becomes popular nowadays are micro aerial vehicles (MAVs) by their advantage to fly in closed and reduced spaces.

Robust guidance, navigation, and control systems for MAVs [1] depend on the input information obtained from on-board sensors as cameras. Undesired movements are usually generated during the fly as a result of complex aerodynamic characteristics of the UAV. Unnecessary image rotations and translations appear in the video sequence, increasing the difficulty to control the vehicle.

There are multiple techniques in the literature [2-5] designed to compensate the effects of undesired movements of the camera. Recently, the video stabilization algorithm 'L1 Optimal' provided by the YouTube editor was introduced in [6]. Another interesting proposal is the Parrot's Director Mode, implemented as an iOS application (iPhone operative system) for post-processing of videos captured with Parrot's AR.Drones.

Usually, offline video stabilization techniques are divided in three stages:

- Local motion estimation
- Motion intention estimation
- Motion compensation

### Local motion estimation

In this phase, the parameters that relate the uncompensated image and the image defined as reference are determined frame by frame. Optical flow [7,8] and geometric transformation models [9-11] are two common approaches for local motion estimation. Our algorithm uses the latter one.

Geometric transformation models are based on the estimation of the motion parameters. For this estimation, interest points should be detected and described. A list of techniques performing this task can be found in the literature [12-14], but Binary Robust Invariant Scalable Keypoints (BRISK) [15], Fast Retina Keypoint (FREAK) [16], Oriented FAST and Rotated BRIEF (ORB) [17], Scale Invariant Feature Transform [18] (SIFT), and Speeded Up Robust Feature (SURF) [19] are common in solving computer vision problems [20]. We are using SURF in this phase as a state-of-the-art algorithm because our contribution is not focused on reducing delays due to the calculation of interest points, not being significant. The delay due to smoothing techniques is higher.

\*Correspondence: wilbert.aguilar@upc.edu  
Automatic Control Department, UPC-BarcelonaTech, Pau Gargallo Street 5,  
08028 Barcelona, Spain

The second part of the motion estimation process is interest points matching across consecutive frames. This is a critical part because the estimated motion parameters are directly dependent on the reliability of matched points. False correspondences will be removed using an iterative technique called Random Sample Consensus (RANSAC), a widely used technique based on a model from a set of points [21-24]. In our work, RANSAC uses a simple cost function based on gray level difference, minimizing the delay.

### Motion intention estimation

In a second phase, for ensuring coherence in the complete motion sequence, the parameters estimated previously are validated in the global and not just in the relative motion between consecutive frames. The main objective of motion intention estimation is to obtain the desired motion in the video sequence suppressing high-frequency jitters from the accumulative global motion estimation.

Several motion smoothing methods are available for motion intention estimation such as particle filter [10], Kalman filter [11], Gaussian filter [25,26], adaptive filter [26,27], spline smoothing [28,29], or point feature trajectory smoothing [30,31]. In our approach, the control signal sent to the MAV is dealt as a known information; hence, a new and different methodology to those in the literature is considered. A combination of a second-order low-pass filter, using as few frames as possible, and action control input is employed to estimate a reliable motion intention. We achieve to reduce the number of frames (time window) required for the smoothing signal using an optimization process.

Most of the techniques cited perform well in video stabilization applications, but there is an additional challenge, not studied in the literature and introduced in this paper, which we have called 'phantom movement'.

### Phantom movement

A phantom movement is mainly a false displacement generated in the scale and/or translation parameters due to the compensation of the high-frequency movements in the motion smoothing. Sometimes the motion smoothing process removes real movements and/or introduces a delay in them. Both cases are defined as phantom movements. This phenomenon represents a problem when teleoperating the MAV, and its effects in other state-of-the-art algorithms will be shown in the 'Results and discussion' section.

Additionally, our proposal to solve this problem will be explained in the 'Real-time video stabilization' section.

### Motion compensation

Finally, the current frame is warped using parameters obtained from the previous estimation phase to generate a stable video sequence.

This paper is organized as follows: the estimation of local motion parameters in our method is explained in the next section. In addition, we describe the combination of RANSAC and gray level difference-based cost function for robust local motion estimation. In the 'Motion intention estimation' section, we present a motion smoothing based on a low-pass filter. Then, the 'Real-time video stabilization' section focuses on the optimization of the algorithm with minimum number of frames to estimate the motion intention. Furthermore, we propose a novel approach to solve the problem of phantom movement. Experimental results and conclusions are presented in the last section.

### Robust local motion estimation

To obtain a stabilized video sequence, we estimate the inter-frame geometric transformation, i.e., the local motion parameters. In this phase, we determine the relationship between the current and the reference frame as a mathematical model. This process can be structurally divided into two parts: (a) interest point detection, description, and matching and (b) inter-frame geometric transformation estimation using matched points. Additionally, an extra process to ensure robustness is (c) robust cumulative motion parameters.

### Interest point detection, description, and matching

As mentioned in the latter section, there are several techniques for detecting and matching interest points. According to the results presented in [20], the computational cost of SURF is considerably lower than that of SIFT, with equivalent performance. Using the Hessian matrix and a space-scale function [32], SURF locates way-points and describes their features using a 64-dimensional vector. Once the vector descriptors are obtained, the interest point matching process is based on minimum Euclidian distance in the 64-dimensional feature space.

### Inter-frame geometric transform estimation using matched points

From the matched interest points, motion parameters between the current and the reference frame can be estimated. Variations between two specific frames are mathematically expressed by the geometric transformation which relates feature points in the first frame with their correspondences in the second frame [33-35],

$$\mathbf{I}_{sp} = \mathbf{H}_t \cdot \mathbf{I}_t \quad (1)$$

where  $\mathbf{I}_{sp} = [x_{sp}, y_{sp}, 1]^T$  and  $\mathbf{I}_t = [x_t, y_t, 1]^T$  are the coordinates of the interest points at the reference image

and the uncompensated image, respectively, and  $\mathbf{H}_t$  is the  $3 \times 3$  geometric transformation matrix. How to choose the reference image will be analyzed in the next subsection.

The geometric transformation represented by  $\mathbf{H}_t$  can be characterized by different models as appropriate. The most common models are translation, affine, and projective models.

In previous works [36,37], we have found experimentally that in both cases, handheld devices and on-board cameras of flying robots, most of the undesired movements and parasitic vibrations in the image are considered significant only on the plane perpendicular to the roll axis (scale, rotation, and translations). This type of distortion can be modeled by a projective transformation, and we use the affine model in our algorithm as a particular case of the projective one [38]. The benefit is twofold: a lower computation time than for the projective model and its ability for direct extraction of relevant motion parameters (scale, rotation roll, and translations in the  $xy$ -plane).

The affine model will determine different roll angles for  $\mathbf{H}_t(1, 1)$ ,  $\mathbf{H}_t(1, 2)$ ,  $\mathbf{H}_t(2, 1)$ , and  $\mathbf{H}_t(2, 2)$ . However, we estimate the mean angle adjustable to these values. This model is called nonreflective similarity and is a particular case of the affine model,

$$\mathbf{H}_t = \begin{bmatrix} s \cos(\phi) & -s \sin(\phi) & t_x \\ s \sin(\phi) & s \cos(\phi) & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

### Robust cumulative motion parameters

Robustness in our algorithm directly depends on the correct matching of interest points in consecutive frames. RANSAC (see Algorithm 1) is a reliable iterative technique for outlier rejection on a mathematical model, in this case, the affine model.

---

#### Algorithm 1 RANSAC algorithm based on cost function $\sum_j J_j$

---

- 1: **for**  $j = 1$  to  $N$  **do**
  - 2:  $j$ th affine transform estimation:  $\mathbf{H}_j$
  - 3:  $j$ th warping of the  $i$ th frame:  $\mathbf{Frame}'_j$
  - 4:  $j$ th cost function computation:  $J_j = \sum_j \left| \mathbf{Frame}'_j - \mathbf{Frame}_{sp} \right|$
  - 5: **end for**
  - 6: Selection of parameters of  $\mathbf{H}_{opt}$  for cost function minimization:  $\arg \min_{(\phi, s, t_x, t_y)} \sum_j \left| \mathbf{Frame}'_j - \mathbf{Frame}_{sp} \right|$
- 

Affine transform can be estimated from three pairs of noncollinear points, but SURF and other techniques obtain hundreds of these pairs of points. RANSAC is performed iteratively  $N$  times using three pairs for each  $\mathbf{H}_j$ ,

obtaining  $N$  different affine transforms. The value of  $N$  is based on the required speed of the algorithm. An alternative is to use an accuracy threshold, but this procedure can be slower.

The cost function of RANSAC algorithm is a key point. In our proposal, we use the absolute intensity difference, pixel by pixel, between the warped and the reference frame. The intensity on a pixel can be affected by common problems such as lighting changes; however, this is not significant in consecutive frames. Therefore, the parameters of the affine model  $\mathbf{H}_{opt}$  that minimize the cost function are

$$\arg \min_{(\phi, s, t_x, t_y)} \sum_j \left| \mathbf{Frame}'_j - \mathbf{Frame}_{sp} \right| \quad (3)$$

The affine model has been selected as the geometric transform between two frames, the one to be compensated and another used as reference, and there are several alternatives for selecting the reference frame. An experimental comparative study was carried out in [37] on three candidates to be the reference frame: the initial frame ( $\mathbf{Frame}_{sp} = \mathbf{Frame}_0$ ), the previous frame ( $\mathbf{Frame}_{sp} = \mathbf{Frame}_{i-1}$ ), and the compensated previous frame ( $\mathbf{Frame}_{sp} = \mathbf{Frame}'_{i-1}$ ). The analysis for the three proposed approaches was performed by using data obtained from an on-board camera of a real micro aerial vehicle. The obtained results show that the approach based on the previous frame is the best candidate to reference.

Finally, the transformation matrix  $\mathbf{H}_{opt}$  is calculated and applied on the current frame for obtaining a warped frame similar to the reference frame, i.e., a stable video sequence.

### Motion intention estimation

The RANSAC algorithm, based on the minimization of the gray level difference, is enough for obtaining a high accuracy in the image compensation of static scenes (scenes without moving objects) [36,37]. However, our goal is to achieve a robust stabilization of video sequences obtained with on-board cameras in micro-aerial vehicles. Most of the unstable videos captured with either flying robots or handheld devices contain dynamic scenes (scenes with moving objects) mainly due to the camera motion. In this way, some movements of the capture device should not be eliminated, but softly compensated, generating a stable motion video instead of a static scene.

The process of approximating the capture device's movements is known as motion intention. Several video stabilization algorithms use smoothing methods for the motion intention estimation such as Kalman filter, Gaussian filter, and particle filter. Our approach is based on a second-order Butterworth filter, a low-pass filter used for smoothing of signals [39].

Our experimentation platform is a low-cost MAVs, which shows a complex dynamic behavior during indoor flight. Consequently, the videos captured with on-board cameras usually contain significant displacements and high speed movements on the plane perpendicular to the roll axis. Effects of wireless communication problems such as frame-by-frame display, low-frequency videos, and video freezing should be also considered.

Using the low-pass filter as a motion intention estimator, most of the problems associated to indoor flight are avoided. However, freezing effects can eventually still appear by a low communication quality. Motion parameters computed from frozen frames must be discarded before to continue with the estimation process.

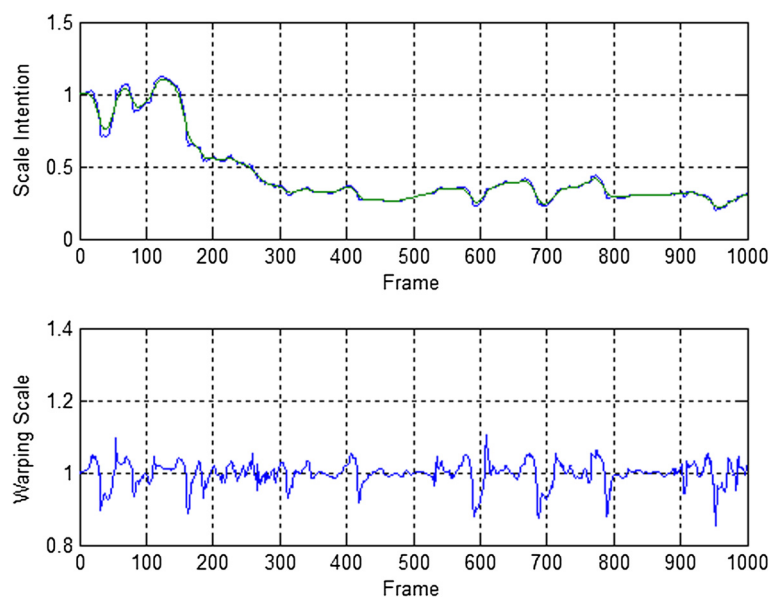
Once the affine transformation parameters (scale, rotation, and translations  $x$  and  $y$ ) are extracted, as well as the values of parameters from frozen screens are removed, the low-pass filter computes the motion intention as an output without high-frequency signals. Low frequencies are associated to the intentional motion, and high frequencies are referred to undesired movements, thus the cutting frequency depending on the application and system characteristics. For cutting frequency, a higher value means an output video similar to the original movements, including the undesired movements, while a lower value means that the output video eliminates intentional movements. In our case, we use a second-order filter with the same cutting frequency 66.67 Hz to smooth the signals of the four motion parameters. An

alternative option is to use a different filter for each motion parameter.

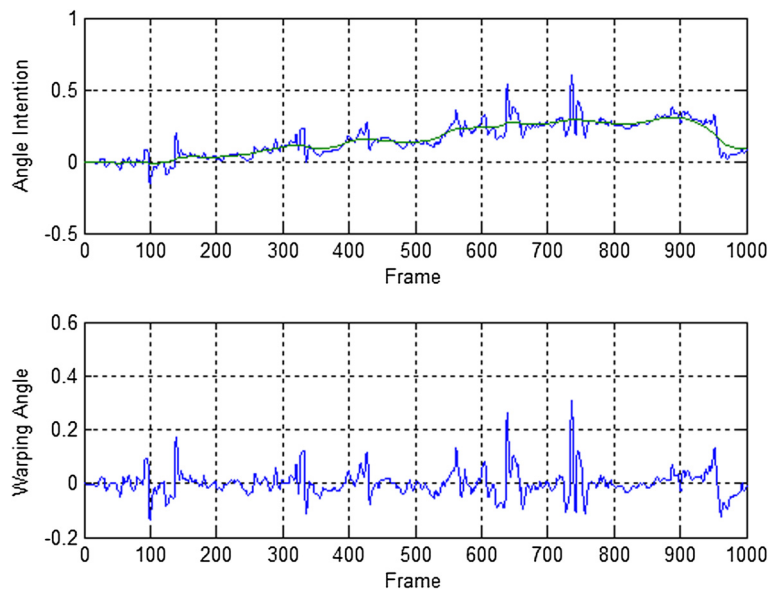
The undesired movement can be estimated by the subtraction of the motion intention, obtaining a high-frequency signal. This signal is then used in image warping to compensate vibrations and, simultaneously, to keep intentional motions. It can be seen in Figure 1 the motion intention signal estimated with the low-pass filter (top) and the high-frequency signal (warping parameter in the figure) to be compensated (down) for the parameter angle. Similar graphics can be obtained for the scale and translations in the  $x$ -axis and  $y$ -axis (Figures 2, 3, and 4).

### Real-time video stabilization

A robust post-processing algorithm for video stabilization has been detailed; however, the goal is a real-time version. In this context, it is worth noting that there are very few techniques for real-time video stabilization, and the first challenge to be solved being computational cost. Hence, calculation time is minimized in [40] by using efficient algorithms of interest point detection and description. This method reduces time in motion intention estimation by means of a Gaussian filter without accumulative global motion, using the stabilized frames in addition to the original frames. Our proposal uses an off-line optimization process for obtaining the minimum number of frames that can be applied in real time to the system at hands without decrease in initial off-line video stabilization performance. Furthermore, this filter will be combined with the known



**Figure 1 Angle.** Top: accumulative (blue) and intentional (green) motion signals estimated with the low-pass filter. Bottom: high-frequency signal to be compensated.



**Figure 2 Scale.** Top: accumulative (blue) and intentional (green) motion signals estimated with the low-pass filter. Down: high-frequency signal to be compensated.

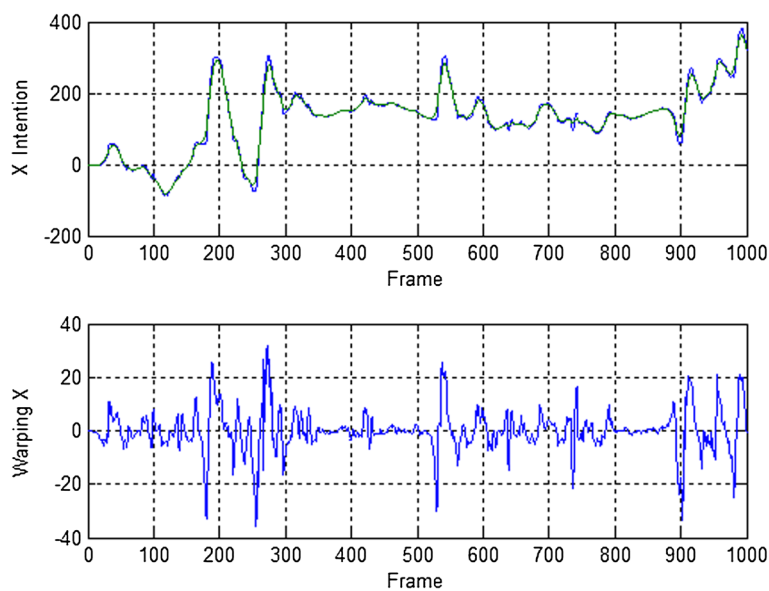
control action signal in order to eliminate the so-called ‘phantom’ movements (in fact, a sort of ‘freezing’) in the compensated video.

### Optimized motion intention estimation

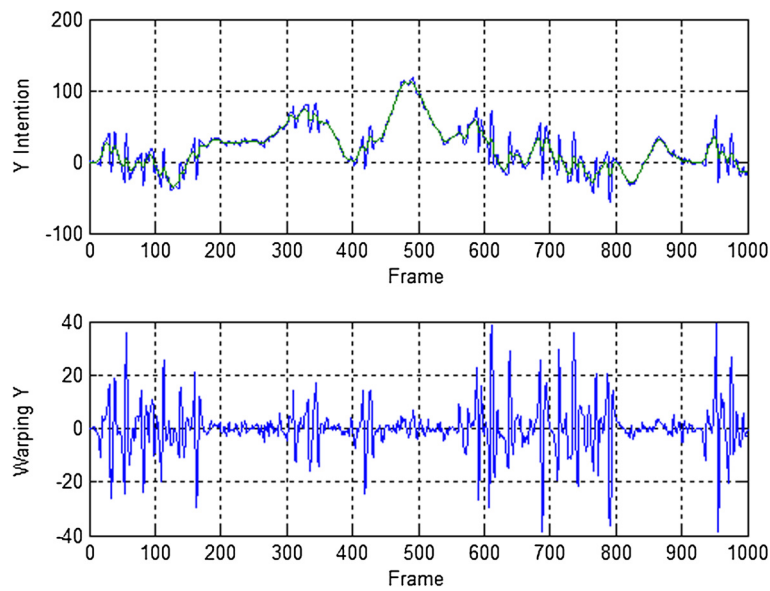
To minimize the number of frames required in the video stabilization process, an exhaustive search has been

implemented by an algorithm that iteratively increases the number of frames used to estimate the motion intention, whose results are plotted in Figure 5.

For the optimization process, it is necessary to define an evaluation metric of the video stabilization performance. Subjective evaluation metrics can be found in the literature, such as the mean opinion score (MOS), which



**Figure 3 Translation in the  $x$ -axis.** Top: accumulative (blue) and intentional (green) motion signals estimated with the low-pass filter. Bottom: high-frequency signal to be compensated.

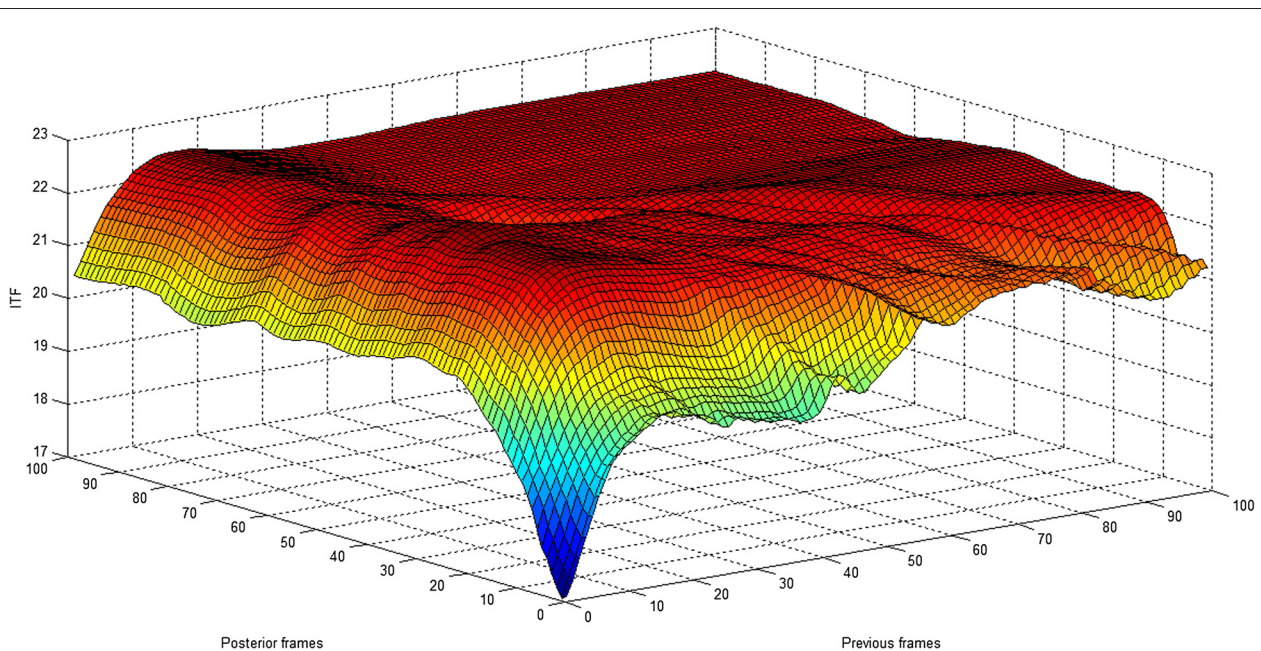


**Figure 4 Translation in the y-axis.** Top: accumulative (blue) and intentional (green) motion signals estimated with the low-pass filter. Bottom: high-frequency signal to be compensated.

is very common in the quality evaluation of the compressed multimedia [41]. The other possibility is to use objective evaluation metrics such as bounding boxes, referencing lines, or synthetic sequences [42]. The advantage of the three referred objective metrics is that estimated motion parameters can be directly compared against real motion. The inter-frame transformation fidelity (ITF) [40]

is a widely used method to measure the effectiveness and performance of video stabilization, whose mathematics expression is

$$ITF = \frac{1}{N_f - 1} \sum_{k=1}^{N_f-1} PSNR(k) \quad (4)$$



**Figure 5 Minimization of inter-frame transformation fidelity (ITF).** We obtain, using four (4) previous and four (4) posterior frames or using only six (6) previous frames, a performance as good as using the complete video sequence to estimate the motion intention.

where  $N_f$  is the number of video frames and

$$\text{PSNR}(k) = 10 \log_{10} \frac{I_{\text{P}_{\text{MAX}}}}{\text{MSE}(k)} \quad (5)$$

is the peak signal-to-noise ratio between two consecutive frames, with

$$\text{MSE}(k) = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|\mathbf{Frame}_k(i, j) - \mathbf{Frame}_{k-1}(i, j)\|^2 \quad (6)$$

being the mean square error between monochromatic images with size  $M \cdot N$  and  $I_{\text{P}_{\text{MAX}}}$  the maximum pixel intensity in the frame.

Based on the optimization of the objective evaluation metric ITF, two solutions have been obtained with a performance as high as using the complete video sequence to estimate the motion intention: a) using four (4) previous and four (4) posterior frames and b) using only six (6) previous frames.

Our work is focused on the real-time application, so it is important to analyze how this issue is affected for both options: a) For the first case, to use four previous and four posteriors frames means that the algorithm will be launched four frames after the video sequence initialization, and the stabilized sequence will be ready after four frames. b) In the second case, the algorithm starts six frames after the video sequence initialization, two frame later than in the first case, but for the rest of the sequence, it can be applied without added delay. Considering that the sample frequency was 10 Hz for the system at hands, a 0.4-s delay would be introduced in the total computational time when using the first option. Our algorithm uses the second option, which relies only on precedent information.

### Phantom movements

Previous video stabilization approaches have obtained good results eliminating undesired movements in images captured with handheld devices and complex systems, but all of them were evaluated using the ITF as cost function. Although the final video has achieved a good ITF performance, i.e., a stable video, the motion smoothing process has generated phantom movements.

For video post-processing applications, the main objective is to stabilize the video; hence, phantom movements do not represent a problem. Notwithstanding, for real-time applications, the objective is to obtain a stable video, but it should be as real as possible. In this sense, it is important to decrease the difference between the real and estimated motion intention, preserving the ITF performance.

The root mean square error (RMSE) [43] is adopted in order to evaluate the reliability of the estimated motion

with respect to the observed motion. A low RMSE means that the estimated motion intention is similar to the real motion intention.

Consequently, the proposed objective evaluation metric to be optimized is the difference between the estimated global motion and the observed motion, measured as RMSE:

$$\text{RMSE} = \frac{1}{2F} \left( \sqrt{\sum_{j=0}^F (E_{x,j} - T_{x,j})^2} + \sqrt{\sum_{i=0}^F (E_{y,i} - T_{y,i})^2} \right) \quad (7)$$

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

Two alternatives exist to merge information when computing real motion intention ( $T_{x,j}$ ,  $T_{y,i}$ ): information obtained from the on-board inertial measurement unit (IMU) or the control action data. Choice depends on the accuracy of the model. In our algorithm, the control action is employed since IMU information is not very reliable in most of the micro aerial vehicles. In this way, the observed motion is defined as a combination between the control action and the smoothed motion signal.

Our algorithm of motion intention estimation (see Algorithm 2) uses control action as a logical gate allowing the execution of the low-pass filter only when a teleoperated motion intention is present. Additionally, our algorithm inserts a hysteresis after the execution of the action control. The objective of this hysteresis is that the system reaches its maximal (or minimal, according to the control action signal) position before the effect of a new control action.

---

### Algorithm 2 Algorithm of reduction of phantom movements using the control action information

---

- 1:  $\{U_i$  is the current control action,  $E_{R_i}$  is the estimated current motion without phantom movements,  $E_i$  is the estimated current motion using the filter}
  - 2: **if**  $U_i \neq 0$  **then**
  - 3:      $E_{R_i} = E_i$
  - 4: **else if**  $((U_{i-1} > 0) \wedge (E_i > E_{i-1}) \vee (U_{i-1} < 0) \wedge (E_i < E_{i-1}))$  **then**
  - 5:      $E_{R_i} = E_i$
  - 6: **else**
  - 7:      $E_{R_i} = R_F \cdot (E_i - E_{i-1}) + E_{i-1}$
  - 8: **end if**
-

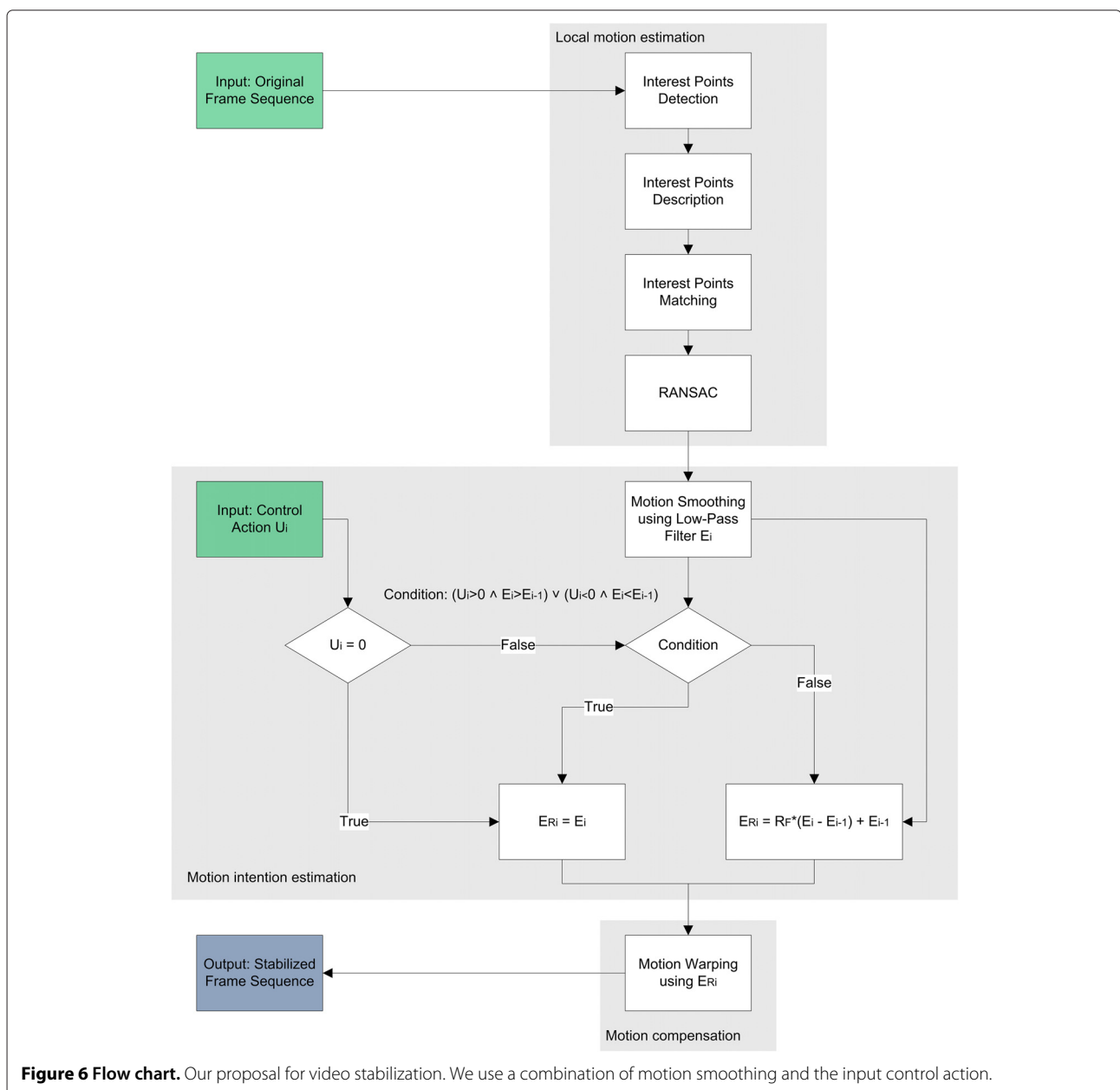
We have defined a reliability parameter  $0 < R_F < 1$ . A value of  $R_F$  close to one leads to achieve a higher ITF value, i.e., a more stable video with phantom movements. On the other hand, using a value of  $R_F$  close to zero, we obtain a less stable video without phantom movements. Our complete algorithm is shown in Figure 6.

### Results and discussion

This section has been divided into three parts: experimental design, video stabilization performance, and comparison with another algorithm.

### The experimental design

The AR.Drone 1.0, a low-cost quadrotor built by the French company Parrot (Paris, France), has been used as experimental platform for several reasons: low cost, energy conservation, safe flight, and vehicle size. The proposed methodology has been implemented in a laptop with the following characteristics: Intel Core i7-2670QM processor, 2.20 GHz with Turbo Boost up to 3.1 GHz and RAM 16.0 Gb. Real images of four different scenarios are obtained with the on-board camera (sample frequency = 10 Hz) and processed. Furthermore, a video has been recorded with a zenith camera to capture the real motion



**Figure 6 Flow chart.** Our proposal for video stabilization. We use a combination of motion smoothing and the input control action.





**Figure 7 Scene 1.** Top: original video. Bottom: stabilized video.

of the flying robot in the  $xy$ -plane. RMSE is selected as objective measure of motion reliability, comparing the estimated motion with the observed motion. In order to obtain a position measure, we use a tracker based on optical flow [44] and camera calibration method for radial distortion [45]. Next, the RMSE is computed by comparing the estimate with the observed motion (from the zenith camera).

#### Video stabilization performance

Now, a visual perception of the results obtained for each experimental environment is shown<sup>a</sup> in Figures 7, 8, 9, and 10. Experiments demonstrate that the approach based on motion intention estimation presented in this paper is robust to the presence of nearby objects, scenes with

moving objects, and common problems described in past sections from on-board cameras for MAVs during indoor flight.

#### Presence of nearby objects

Nearby objects in the scene represents one of the main problems of video stabilization, because most of interest points are generated in the objects' region. The image compensation is computed using the objects' motion instead of the scene motion. However, our process of matching interest points is based on the RANSAC algorithm and the gray level difference between consecutive frames as cost function. Consequently, the process of motion estimation is not performed on the objects' interest points but on the whole scene.



**Figure 8 Scene 2.** Top: original video. Bottom: stabilized video.



**Figure 9 Scene 3.** Top: original video. Bottom: stabilized video.

### **Scenes with moving objects**

Moving objects are another common problem. Some objects with many points cause, during the motion estimation, undesirable tracking of these objects. Once more, the RANSAC-based process of matching interest point is not only referenced to moving objects but to the whole image.

### **Problems from on-board cameras for quadrotors**

Scenes frame by frame, significant displacements, low-frequency videos, freezing, and high-speed displacements are frequent problems in images captured with an on-board camera due to the complex dynamic of the quadrotors during indoor flight. In all of them, the change between two consecutive frames could be considerable, producing a critical problem in video stabilization. In

our approach, motion intention estimation solves these problems, and previous rejection of data higher than a threshold provides additional robustness.

### **Phantom movements**

They corresponds to a phenomenon present in previous video stabilization techniques, but not still reported. Independently of which approach is used, video stabilization process depends on a phase of motion intention estimation. The phantom movements are generated during the elimination of the high-frequency movements due to the motion intention estimation which reduces the frequency of the movements and the previous motion intention estimators which are not able to detect or correct these troubles. Our proposal eliminates this phantom movements using a combination of a low-pass filter, as a motion



**Figure 10 Scene 4.** Top: original video. Bottom: stabilized video.

**Table 1 Evaluation metrics**

Video name	Evaluation metric	Original	L1-Optimal	Our approach
Video 1	ITF (dB)	14.09	19.62	19.48
	RMSE		0.046	0.028
Video 2	ITF (dB)	13.43	19.57	19.52
	RMSE		0.051	0.023
Video 3	ITF (dB)	14.65	20.16	19.89
	RMSE		0.047	0.029
Video 4	ITF (dB)	16.96	21.24	21.12
	RMSE		0.036	0.017

ITF is the measure of video stability, RMSE is the measure of reliability of the video movements.

intention estimator, with the control action. However, this method slightly decreases the ITF value.

### Comparison

Our approach has been compared with the off-line method L1-Optimal [6], which is applied in the YouTube Editor as a video stabilization option. Results on four different scenes are presented in Table 1 using two evaluation metrics: ITF and RMSE.

The obtained results show that our algorithm is comparable with the L1-Optimal method. The performance of our approach with respect to the ITF measure is slightly lower than L1-Optimal, which means that the output video is a little less stable. However, the RMSE value, which represents the similarity between the stabilized

video motion and the intentional motion, is higher in our approach.

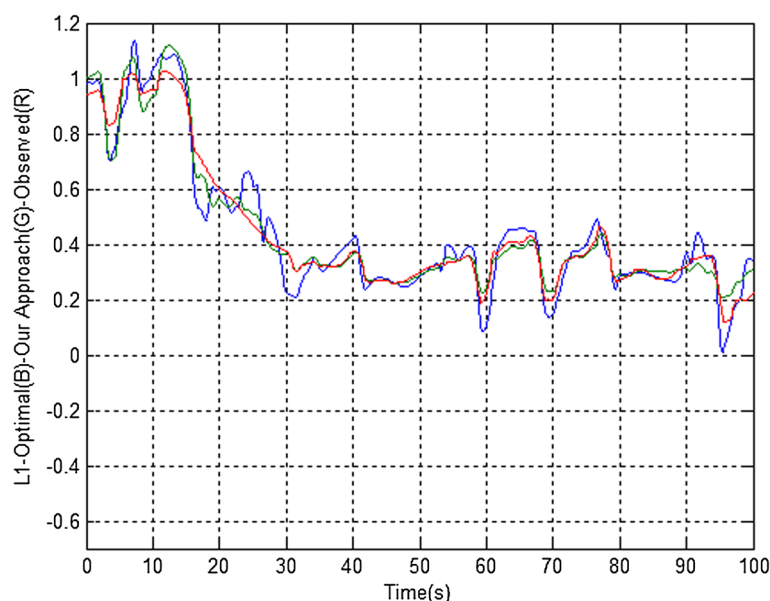
It is worth noting that the ITF measure could be increased varying the reliability factor  $R_F \in [0, 1]$ , but the RMSE inevitably would decrease. For post-production applications, the value for  $R_F$  can be equal to one, but for applications of motion control based on camera information, the realism of the movement is important, so the reliability factor should decrease to zero. Both, observed and estimated scales are graphically compared in Figure 11 for the technique L1-Optimal and our approach.

### Conclusions

After conducting an initial study of motion smoothing methods, it has been experimentally checked that the low-pass filter has a high performance as algorithm for motion intention estimation, eliminating undesired movements.

However, this method can be optimized using a lower number of frames without decreasing the ITF measure, as we have presented in this paper. The cutting frequency depends on the model characteristics; hence, the information about the capture system implicates a considerable contribution in a calibration phase.

The phantom movements are a phenomenon that had not yet been studied in the video stabilization literature, but it is a key point in the control of complex dynamic systems as micro aerial vehicles, where the realism in the movements could mean the difference that prevents an accident.



**Figure 11 Comparison of the scales.** L1-Optimal (blue), our approach (green), and observed (red).

The reliability factor is adapted to the purpose of the application. This application can be a post-production with a high ITF value and a lower realism or the opposite situation, for real-time video stabilization used in tele-operation systems.

As a future work, we will extend our video stabilization method, using the quadrotor model estimated in [46], for aggressive environments with turbulence and communication problems. We will apply it for increasing the performance of detection and tracking algorithms. In [47] we presented a first application for face detection.

## Endnote

<sup>a</sup>Video results are provided: [www.youtube.com/user/VideoStabilizerMAV/videos](http://www.youtube.com/user/VideoStabilizerMAV/videos).

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness, through the PATRICIA project (TIN 2012-38416-C03-01). The research fellow Wilbert G. Aguilar thanks the funding through a grant from the program 'Convocatoria Abierta 2011' issued by the Secretary of Education, Science, Technology and Innovation SENESCYT of the Republic of Ecuador.

Received: 17 December 2013 Accepted: 5 September 2014

Published: 22 September 2014

## References

1. F Kendoul, Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems. *J. Field Robot.* **29**(2), 315–378 (2012)
2. Z Duric, A Rosenfeld, Shooting a smooth video with a shaky camera. *Mach. Vis. Appl.* **13**(5–6), 303–313 (2003)
3. S Battiato, G Gallo, G Puglisi, S Scellato, SIFT features tracking for video stabilization, in *14th International Conference on Image Analysis and Processing, 2007. ICIAP 2007* (Modena, 10–14 Sept 2007), pp. 825–830
4. Y-F Hsu, C-C Chou, M-Y Shih, Moving camera video stabilization using homography consistency, in *2012 19th IEEE International Conference on Image Processing (ICIP)* (Lake Buena Vista, 30 Sept–3 Oct 2012), pp. 2761–2764
5. C Song, H Zhao, W Jing, H Zhu, Robust video stabilization based on particle filtering with weighted feature points. *Consum. Electron. IEEE Trans.* **58**(2), 570–577 (2012)
6. M Grundmann, V Kwatra, I Essa, Auto-directed video stabilization with robust l1 optimal camera paths, in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Colorado Springs, 20–25 June 2011), pp. 225–232
7. H-C Chang, S-H Lai, K-R Lu, A robust and efficient video stabilization algorithm, in *2004 IEEE International Conference on Multimedia and Expo, 2004. ICME '04*, vol. 1 (Taipei, 27–30 June 2004), pp. 29–321
8. R Strzodka, C Garbe, Real-time motion estimation and visualization on graphics cards. *IEEE Vis.* **2004**, 545–552 (2004)
9. K-Y Lee, Y-Y Chuang, B-Y Chen, M Ouhyoung, Video stabilization using robust feature trajectories, in *2009 IEEE 12th International Conference on Computer Vision* (Kyoto, 27 Sept–4 Oct 2009), pp. 1397–1404
10. J Yang, D Schonfeld, M Mohamed, Robust video stabilization based on particle filter tracking of projected camera motion. *Circuits Syst. Video Technol. IEEE Trans.* **19**(7), 945–954 (2009)
11. C Wang, J-H Kim, K-Y Byun, J Ni, S-J Ko, Robust digital image stabilization using the Kalman filter. *Consum. Electron. IEEE Trans.* **55**(1), 6–14 (2009)
12. J Canny, A computational approach to edge detection. *Pattern Anal. Mach. Intell. IEEE Trans. PAMI-8(6), 679–698 (1986)*
13. C Harris, M Stephens, A combined corner and edge detector, in *Proceedings of The Fourth Alvey Vision Conference*, (31 Aug–2 Sept 1988), pp. 147–151
14. O Miksik, K Mikolajczyk, Evaluation of local detectors and descriptors for fast feature matching, in *2012 21st International Conference on Pattern Recognition (ICPR)* (Tsukuba, 11–15 Nov 2012), pp. 2681–2684
15. S Leutenegger, M Chli, RY Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, in *2011 IEEE International Conference on Computer Vision (ICCV)* (Barcelona, 6–13 Nov 2011), pp. 2548–2555
16. A Alahi, R Ortiz, P Vanderghenst, Freak: Fast retina keypoint, in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Providence, 16–21 June 2012), pp. 510–517
17. E Rublee, V Rabaud, K Konolige, G Bradski, ORB: an efficient alternative to SIFT or SURF, in *2011 IEEE International Conference on Computer Vision (ICCV)* (Barcelona, 6–13 Nov 2011), pp. 2564–2571
18. DG Lowe, Object recognition from local scale-invariant features, in *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, vol. 2 (Kerkyra, 20–25 Sept 1999), pp. 1150–11572
19. H Bay, T Tuytelaars, L Gool, SURF: Speeded Up Robust Features, in *Computer Vision – ECCV 2006*, ed. by A Leonardis, H Bischof, A Pinz, and (eds.) Lecture Notes in Computer Science, vol. 3951 (Springer Berlin, 2006), pp. 404–417
20. J Luo, G Oubong, A Comparison of SIFT, PCA-SIFT and SURF. *Int. J. Image Process. (IJIP)*, **3**(4), 143–152 (2009)
21. MA Fischler, RC Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM.* **24**(6), 381–395 (1981)
22. B Tordoff, D Murray, Guided sampling and consensus for motion estimation, in *Computer Vision – ECCV 2002*, ed. by A Heyden, G Sparr, M Nielsen, P Johansen, and (eds.) Lecture Notes in Computer Science, vol. 2350 (Springer Berlin, 2002), pp. 82–96
23. S Choi, T Kim, W Yu, Performance evaluation of RANSAC family, in *Proceedings of the British Machine Vision Conference 2009* (London, 7–10 Sept 2009), pp. 81–18112
24. KG Derpanis, *Overview of the RANSAC algorithm*. (Technical report, Computer Science, York University, 2010)
25. Y Matsushita, E Ofek, W Ge, X Tang, H-Y Shum, Full-frame video stabilization with motion inpainting. *Pattern Anal. Mach. Intell. IEEE Trans.* **28**(7), 1150–1163 (2006)
26. P Rawat, J Singhai, Adaptive motion smoothing for video stabilization. *Int. J. Comput.* **72**(20), 14–20 (2013)
27. S Wu, DC Zhang, Y Zhang, J Basso, M Melle, Adaptive smoothing in real-time image stabilization, in *Visual Information Processing XXI* (Baltimore, 24–25 April 2012)
28. Y Wang, Z Hou, K Leman, R Chang, Real-time video stabilization for unmanned aerial vehicles, in *MVA2011 IAPR Conference on Machine Vision Applications* (Nara, 13–15 June 2011), pp. 336–339
29. Y Wang, R Chang, T Chua, Video stabilization based on high degree b-spline smoothing, in *2012 21st Conference on Pattern Recognition (ICPR)* (Tsukuba, 11–15 Nov 2012), pp. 3152–3155
30. YG Ryu, HC Roh, M-J Chung, Long-time video stabilization using point-feature trajectory smoothing, in *2011 IEEE International Conference on Consumer Electronics (ICCE)* (Las Vegas, 9–12 Jan 2011), pp. 189–190
31. D Jing, X Yang, Real-time video stabilization based on smoothing feature trajectories. *Appl. Mech. Mater.* **519–520**(1662–7482), 640–643 (2014)
32. K Mikolajczyk, C Schmid, Scale & affine invariant interest point detectors. *Int. J. Comput. Vis.* **60**(1), 63–86 (2004)
33. O Faugeras, Q-T Luong, T Papadopoulos, *The Geometry of Multiple Images: the Laws That Govern the Formation of Images of a Scene and Some of Their Applications*. (MIT Press, Cambridge, 2001)
34. DA Forsyth, J Ponce, *Computer Vision: a Modern Approach*. (Prentice Hall, Upper Saddle River, 2002)
35. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, New York, 2003)
36. WG Aguilar, C Angulo, Estabilización robusta de vídeo basada en diferencia de nivel de gris, in *Proceedings of the 8th Congress of Science and Technology ESPE 2013*, (Sangolquí, 5–7 June 2013)
37. WG Aguilar, C Angulo, Robust video stabilization based on motion intention for low-cost micro aerial vehicles, in *2014 11th International Multi-conference on Systems, Signals and Devices (SSD)* (Barcelona, 11–14 Feb 2014), pp. 1–6

38. M Vazquez, C Chang, Real-time video smoothing for small RC helicopters, in *IEEE International Conference on Systems, Man and Cybernetics, 2009. SMC 2009* (San Antonio, 11–14 Oct 2009)
39. SW Bailey, B Bodenheimer, A comparison of motion capture data recorded from a Vicon system and a Microsoft Kinect sensor, in *Proceedings of the ACM Symposium on Applied Perception. SAP '12* (ACM New York, 2012), pp. 121–121
40. J Xu, H-W Chang, S Yang, M Wang, Fast feature-based video stabilization without accumulative global motion estimation. *Consum. Electron. IEEE Trans.* **58**(3), 993–999 (2012)
41. M Niskanen, O Silven, M Tico, Video stabilization performance assessment, in *2006 IEEE International Conference on Multimedia and Expo* (Toronto, 9–12 July 2006), pp. 405–408
42. S-J Kang, T-S Wang, D-H Kim, A Morales, S-J Ko, Video stabilization based on motion segmentation, in *2012 IEEE International Conference on Consumer Electronics (ICCE)*, (2012), pp. 416–417
43. C-L Fang, T-H Tsai, C-H Chang, Video stabilization with local rotational motion model, in *2012 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)* (Kaohsiung, 2–5 Dec 2012), pp. 551–554
44. A Yilmaz, O Javed, M Shah, Object tracking: a survey. *ACM Comput. Surv.* **38**(13) (2006)
45. Z Zhang, A flexible new technique for camera calibration. *Pattern Anal. Mach. Intell. IEEE Trans.* **22**(11), 1330–1334 (2000)
46. WG Aguilar, R Costa, C Angulo, L Molina, Control autónomo de cuadricópteros para seguimiento de trayectorias, in *Proceedings of the 9th Congress of Science and Technology ESPE 2014* (Sangolquí, 28–30 May 2014)
47. WG Aguilar, C Angulo, Estabilización de vídeo en micro vehículos aéreos y su aplicación en la detección de caras, in *Proceedings of the 9th Congress of Science and Technology ESPE 2014* (Sangolquí, 28–30 May 2014)

doi:10.1186/1687-5281-2014-46

**Cite this article as:** Aguilar and Angulo: Real-time video stabilization without phantom movements for micro aerial vehicles. *EURASIP Journal on Image and Video Processing* 2014 **2014**:46.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)