2013 AASRI Conference on Intelligent Systems and Control

Video Stabilization for Aerial Video Surveillance

Ahlem Walha[a], Ali Wali[a] and Adel M. Alimi[a] *

*[a] REGIM-Lab. Research Laboratory on Intelligent Machines,
University of Sfax, National Engineering Scool of Sfax (ENIS), BP1173
Sfax 3038, Tunisia*

Abstract

Aerial video stabilization system aims to remove undesired motion in aerial video. This motion is the result of undesired movement of mobile sensor. In this article we present a new video stabilization system for Unmanned Aerial Vehicles (UAV). Our system is based on keypoints tracking. We use Scale Invariant Feature Transform (SIFT) keypoint detection, and matching to estimate parameters of affine transformation model. Then, Kalman filter with median filter is applied to remove video noise. A number of real aerials videos surveillances demonstrate that this method can achieve good performance.

1. Introduction

The quality of output video in mobile surveillance systems suffers from different undesired jitter like track, unwanted vibrated motion, boom or pan. Video stabilization is used to create a new video sequence where the undesired motion between frames has been removed. Therefore it has become essential in many mobile

* Corresponding author. Tel.: +216-21749702; fax: +216-74677545.
E-mail address: walha.ahlem, ali.wali, adel.alimi@ieee.org

surveillance systems such as unmanned aerial vehicle (UAV) systems. Also it is the first step in many aerial applications such as background estimation and object tracking [1].

Techniques of video stabilization can be divided into four groups: optical approach, mechanical approach, electronic approach and digital approach. In this paper we focused on digital approach. This technique is an image pre-processing. All digital stabilization systems handle three essential aspects. The first one estimates the global motion, the second one concerns the motion smoothing and the last one is related to the motion compensation. Among these components, the step of global motion estimation is the most vital but also the most difficult one [9].

In case of a fixed camera, a strong winds or small vibration from heavy traffic can caused the global motion. In this case background is almost fixed over the long term [15]. Consequently, motion can be calculated by local point tracking [1], or by searching a region that contain few motions [7]. In the case of mobile platform, on which the research is conducted, the global motions include two elements: the intentional and the unwanted motion. Several efforts have been put in case of mobile camera. Block matching techniques improves motion estimation by using different adaptive filters [9]. These method present good results if the video does not contain moving object [8].

In this paper we present a new system to stabilize aerial video surveillance by extracting and matching SIFT point for consecutive frames. In the following, Sec.2 cites some related works on video stabilization; Sec. 3 describes our proposed system. Results achieved by our system are presented in Sec.4. Finally Sec.5 presents summarized conclusions.

2. Related work

The goal of feature based algorithms is to estimate interframe motion by extraction features from video images [15]. Some techniques [6] that combine features extraction with other robust filters have good performances. Motion filtering is also a significant step in video stabilization process. In this step undesired movement is recognized by evaluation of estimated motion. Different methods have been introduced to correct translational and rotational jitters. Kalman filtering [10] and extended Kalman filtering, Frame Position Smoothing [13], Gaussian filtering [17] and Motion Vector Integration [14] are among these techniques. A video stabilization algorithm using SIFT [12] has been introduced in [14]. Junlan et al [14] uses Iterative Least Squares method to reduce estimation error then uses Adaptive Motion Vector Integration to filter intentional camera motion. Another system [19] employs SIFT point in order to calculate interframe motion. They recognize intentional movement by Kalman filtering and reduce error variance by using particle filter. But, in this system original SIFT algorithm's parameters doesn't adapt to video stabilization system. And both Kalman filtering and particle filtering calculated for each frame implies intensive computation.

3. Approach

Our input is a real video captured from UAV. First of all, SIFT point are extracted and matched for two consecutive frame. Next, inter frames motion is estimated using affine transform model. Finally, both Kalman filtering and median filtering are used in the step of frame compensation. The details are explained as follows.

3.1. SIFT Point Extraction and matching

SIFT is presented by David Lowe as a local feature description [12]. SIFT point are invariant to image translation, rotation and scale [2]. Therefore, it can identify and track keypoints over multiple frames of video.

It can also afford robust matching in our case where aerial video challenges are considered such as noise, viewpoint changing and inconstant illumination.

In our approach, we estimate global motion vector by extracting SIFT points from two successive frame. Next we calculate local motion vector by matching they two sets of invariant features. In other words, local motion vector between *frame $_{n-1}$* and *frame $_n$*, can be estimated by extracting both the first and the second keypoints K_{point1} (x_{point1}, y_{point1}, *1)* and K_{point2} (x_{point2}, y_{point2}, *1)* from these two frames. In this step, we can show how the keypoint has probably moved from two successive frames. Then we use RANSAC (Random Sample Consensus) to select optimal matching. But, by this method, we obtain a whole number of local motion vectors. They sets of vector does not contain helpful indication for real movement of the camera because they include matches related to moving objects in the frame. Deal with this problem we assume that, comparing to other motions, the velocity of moving objects in the scene is very large. For this reason we use a fixed threshold to eliminate moving object. As a result, we can generate the transformation matrix.

3.2. Motion Estimation

Motion can be described either by a 2-D model or by a 3-D model [2]. The various transformations occurring in the 2D plane are Translation, Euclidean or rotation, Similarity and Affine. Thus, in our method we adopt a four parameter 2-D affine estimation model to describe geometric transformation between two consecutive frames. Given a point localized as $P_n(x_n, y_n, 1)$ in *frame$_n$*, and located as P_{n+1} ($x_{n+1}, y_{n+1}, 1$) in *frame$_{n+1}$*, the transformation model from P_n to P_{n+1} can be described as:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \\ 1 \end{bmatrix} = A.\begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} \qquad (1)$$

A is an affine matrix precisely described by θ rotation, Sc scaling, and Tr_x and Tr_y translations of the camera in a scene with.

$$A. = \begin{bmatrix} Sc\cos\theta & -Sc\sin\theta & Tr_x \\ Sc\sin\theta & Sc\cos\theta & Tr_y \\ 0 & 0 & 1 \end{bmatrix} \qquad (2)$$

In this matrix has only four free parameters compared to the complete affine transformation matrix which originally has six: one scale, one angle, and two translations. To resolve this issue, we use linear Least Squares Method on a set of iterations. In fact, it can provide robust parameter estimation.

3.3. Motion compensation

In this final step, we need to correct the current frame to obtain stable image. But parameters calculated in equation (1) contain two types of motion: motion of the sensors and normal movement of the UAV. To compensate the current frame we should separate these two types of motion.

Fig. 1 Sample images from VIRAT Aerial Video Dataset

Kalman filter is a basic procedure to put the new frame according to the estimated motion. So we use this filter to estimate the motion of the sensor. Then we use median filter in order to refine the obtained result.

4. Experimental and Results

We test our system on a variety of scenes from VIRAT Aerial Video dataset [16] containing low resolution sequences of 720 x 480 pixels captured in 30 fps. This aerial dataset is characterized by zooming, varying viewpoints and scale. The results of the proposed system are illustrated in Fig. 2

Fig. 2. Original and stabilized video. First row: frame 1,40,80 and 211 of original sequence is shown here. Second row: stabilized sequences.

Peak Signal-to-Noise Ratio (PSNR) is used to evaluate the quality of our stabilized video. PSNR computed between two consecutives frames is defined as:

$$PSNR(n) = 10\log_{10}\frac{I_{MAX}}{MSE(n)} \qquad (3)$$

The PSNR gives a relation between the desired output and the obtained video. In this equation, $MSE(n)$ measure the Mean-Square-Error between successive frames, I_{MAX} is the maximum pixel value of an image. Frame dimensions are represented by *N* and *M.* The PSNR value for each frame of the original video and our stabilized video are shown in Fig 3. Higher PNSR between two stabilized frames represent good quality of stabilized video.

$$MSE(n) = \frac{1}{MN} \sum_{y=1}^{M} \sum_{x=1}^{N} [I_N(x,y) - I_{n+1}(x,y)]^2 \qquad (4)$$
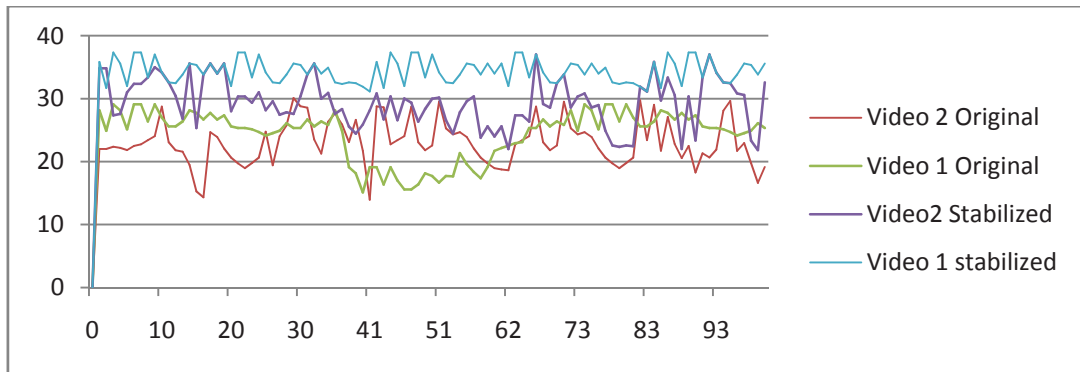

Fig. 3. PSNR of two original video and two stabilized video.

The measurement of Interframe Transformation Fidelity (ITF) is determined by

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

TIF is the average of the PSNR between two consecutives frames. In general this average is used for each value, to obtain an approximate estimation of the quality of the stabilized video. Similar to PSNR, upper ITF values indicate super quality video stabilization. ITF values for three video sequences tested are shown in Table1. This evaluation illustrate that, the ITF of our stabilized videos is superior to the ITF of the original videos. The ITF of our stabilized videos enhances, which is acceptable.

Table. 1 ITF of original and stabilized videos

Videos	Original ITF	Stabilized ITF
Video1	23,74	23,98
Video2	18,04	21,87
Video3	25,65	27,54

5. Conclusions and future works

A new system for aerial video stabilization has been introduced in this article. The main idea of this system is to filtered undesired motion by detecting and matching SIFT point in order to predict the interframe motion. To evaluate our system we used real video captured by a camera installed on UAV. The experimental results prove the efficiency and accuracy of our stabilization system. Our future work will concentrate on performing motion estimation by integrating optical flow in the process of local motion detection.

References

[1] Hong S, Hong T and Wu Y, Multi-resolution unmanned aerial vehicle video stabilization, Aerospace and Electronics Conference (NAECON), pages 126-131, 2010.

[2] Manish O and Prabik. K. B, Improving video stabilization in the presence of motion blur, Computer Vision, Pattern Recognition, Image Processing and Graphics, pages 78-81, 2011.

[3] Rawat P and Singhai J, Review of Motion Estimation and Video Stabilization techniques For hand held mobile video, Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011.

[4] Morimoto C and Chellappa R, Fast electronic digitl image stabilization , Conference on Pattern Recognition, vol. 3, pages .284-288, 1996.

[5] Walha A, Wali A and Alimi AM, Support Vector Machine Approach for Detecting Events in Video Streams, " Advanced Machine Learning Technologies and Applications" 143-151, 2012

[6] Lu W, Hongying Z, Shiyi G, Ying M and Sijie L, The adaptive compensation algorithm for small UAV image stabilization, Geoscience and Remote Sensing Symposium (IGARSS), pages 4391-4394, 2012.

[7] Batur AU and Flinchbaugh B, Video Stabilisation with Optimized Motion Estimation Resolution, International Conference on Image Processing, pages .465-468, 2006

[8] Jesse S J, Zhigang Z, Guangyou X, Digital Video Sequence Stabilization Based on 2.5D Motion Estimation and Inertial Motion Filtering, Real-Time Imaging, Volume 7, Issue 4, Pages 357–365, 2001.

[9] Tico M and Vehvilainen M, Constraint Motion Filtering for Video Stabilisationsing, Proc. of International Conference on Image Processing, pages.569-572, 2005.

[10] Erturk S. Image sequence stabilisation based on kalman filtering of frame positions. Electronics Letters, 37(20), 200.

[12] Lowe D. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, Vol. 60(2), pages 91–110, 2004.

[13] Erturk S, Image sequence stabilisation: motion vector integration (MVI) versus frame position smoothing (FPS), Image and Signal Processing and Analysis, pages 266-271, 2001.

[14] Junlan Y, Schonfeld D and Mohamed M, Robust Video Stabilization Based on Particle Filter Tracking of Projected Camera Motion, Circuits and Systems for Video Technology, IEEE Transactions on , vol.19, no.7, pages 945-954, 2009.

[15] Wali A and Alimi AM, Incremental learning approach for events detection from large video dataset, Advanced Video and Signal Based Surveillance (AVSS), pages 555 – 560, 2010.

[16] Sangmin O, Anthony H, Amitha P, Naresh C, Chia-Chih C, Jong TL, Saurajit M, J. K. A, Hyungtae L, Larry D, Eran S, Xioyang W, Qiang J, Kishore R, Mubarak S, Carl V, Hamed P, Deva R, Jenny Y, Antonio T, Bi S, Anesco F, Amit RC and Mita D. A large-scale benchmark dataset for event recognition in surveillance video. Computer Vision and Pattern Recognition (CVPR), pages 527 – 528, 2011.

[17] Hong S and Atkins E, Moving Sensor Video Image Processing Enhanced with Elimination of Ego Motion by Global Registration and SIFT, Tools with Artificial Intelligence, ICTAI '08, poges37-40, 2008.