Improved image alignment algorithm based on projective invariant for aerial video stabilization

Meng Yi[1,2], Bao-long Guo[1] and Chun-man Yan[3]
[1] Institute of Intelligent Control and Image Engineering, Xidian University
Xi'an 710071, China
[2] School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China
[e-mail: yimeng0120@gmail.com]
[3] College of Physics and Electronic Engineering, Northwest Normal University,
Lanzhou, Gansu 730070, P. R. China
[e-mail: yancha02@163.com]
*Corresponding author: Meng Yi

Received July 16, 2013; revised November 6, 2013; accepted November 6, 2013; published September 30, 2014

Abstract

In many moving object detection problems of an aerial video, accurate and robust stabilization is of critical importance. In this paper, a novel accurate image alignment algorithm for aerial electronic image stabilization (EIS) is described. The feature points are first selected using optimal derivative filters based Harris detector, which can improve differentiation accuracy and obtain the precise coordinates of feature points. Then we choose the Delaunay Triangulation edges to find the matching pairs between feature points in overlapping images. The most "useful" matching points that belong to the background are used to find the global transformation parameters using the projective invariant. Finally, intentional motion of the camera is accumulated for correction by Sage-Husa adaptive filtering. Experiment results illustrate that the proposed algorithm is applied to the aerial captured video sequences with various dynamic scenes for performance demonstrations.

Keywords: Image alignment, Electronic image stabilization, Delaunay triangulation, Projective invariant.

http://dx.doi.org/10.3837/tiis.2014.09.013

1. Introduction

When cameras are mounted on the unstable airplane platforms, it is barely possible to obtain a smooth motion because of undesired camera motions. Electronic image stabilization (EIS) is, therefore, becoming an indispensable technique in advanced digital cameras and camcorders. EIS can be defined as the process of removing unwanted video vibrations and obtaining stabilized image sequences [1-3]. It has been widely used in the areas of video surveillance, panorama stitching, robot localization and moving objects tracking [4-7]. However, making a stable video is a very challenging task especially when an motion of both camera (ego-motion and high-frequency motion) and foreground objects is present,The stabilization accuracy profoundly affects the stabilization quality and impedes the subsequent processes for various applications [8, 9].

The EIS mainly consists of two parts: motion estimation (ME) and motion compensation (MC). The ME is responsible for estimating the reliable global camera movement through three processing steps on the acquired image sequences (see [4] for an alternative scheme): feature detection, feauture matching, and transformation model estimation [10]; the MC can preserve the panning motion of the camera while correcting the undesired fluctuation motions due to an unsteady and vibrating platform. Compared with MC, motion estimation (ME) plays the most important role in EIS and its estimation precision is a decisive step toward video stabilization [11]. In order to enable the use of aerial video in stabilization and reconnaissance missions, the motion estimation (ME) algorithms have to be robustness with respect to different conditions. The block matching algorithm (BMA) [12], bit plane matching (BPM) [13] and phase correlation [14] are the most common ways to stabilize the translational jitter. In this paper, a special class of ME methods is considered that aligns the frames in an aerial video of a dynamic scene captured by a moving camera.

A number of papers have been proposed to realize background motion estimation (ME). The common approaches for background motion estimation include the direct-based methods [15,16] and feature-based methods [17-20]. Direct-based methods aim to find the unknown transforms using raw pixel intensities. Feature-based methods, on the other hand, first identify the feature correspondences between image pairs and then recover the transform considering the correspondence pairs. A method developed by [17] uses the stable relative distance between point sets to delete the local features like the local moving objects, covered points or the inevitable mismatches. [18] uses scale-invariant feature transform (SIFT) [21] points to obtain a crude estimate of the projective transformation, and identifies the features from the moving objects by the difference in moving velocities between objects and the background. [19] achieves ME through detecting SIFT points and calculating the parameters of the projective transformation in a RANSAC process. A Gaussian distribution is used to create a background model and detect the distant moving objects. However, this method only applicable to runway scene. [20] develops a 3D camera motion model, which can be applied to general case.

We propose a method for estimating aerial video motion of a scene consisting of planar background, foreground moving objects and static 3D structures. This system, as shown in **Fig. 1**, aims at presenting a novel optimal derivative filters and projective invariant based ME algorithm, which can generate the accurate locations of the corner points and distinguish more accurate matching points from the less accurate ones, and the most accurate points that belong to the planar background will be used to estimate transformation model. The proposed method

1)selects a set of accurate feature points in each frame based on optimal derivative filters method, 2)finds the matching points between the feature points in the frames, 3)uses the projective invariant method that can distinguish more accurate matching points from the less accurate ones, and the most accurate points that belong to the planar background will be used to estimate transformation model, 4)performs the motion compensation with Sage-Husa Kalman filter [22] to stabilize the sequence.

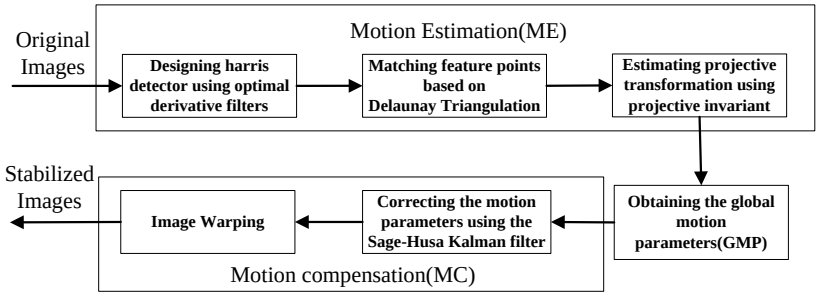


Fig. 1. Flowchart of the proposed aerial video stabilization algorithm.

In the following sections, related work in aerial video stabilization is reviewed, details of an automatic stabilization system are provided, and experimental results of the performance of the algorithm are presented and discussed.

2. Related Work

Image stabilization has been studied extensively over the past decade. In this paper, a special class of video stabilization is taken into account that the camera motion in a video of a dynamic scene is obtained by the moving camera. The challenge of image stabilization in such video is how to track the camera motion accurately without the influence caused by the moving object and static 3-D structures in the images. A number of papers have been proposed to realize background stabilization. The common approaches for background stabilization include the direct-based methods [16, 15] and feature-based methods [18, 19, 20]. Direct-based methods aim to find the unknown transforms using raw pixel intensities. Feature-based methods, on the other hand, first identify the feature correspondences between image pairs and then recover the transform considering the correspondence pairs. A method developed by Zhu J.J et al. [17] uses the stable relative distance between point sets to delete the local features like the local moving objects, covered points or the inevitable mismatches. Yang. J et al. [18] uses scale-invariant feature transform points to obtain a crude estimate of the projective transformation, and identifies the features from the moving objects by the difference in moving velocities between objects and the background. Cheng.H.P et al.[19] achieves aerial video stabilization through detecting SIFT [22] points and calculating the parameters of the projective transformation in a RANSAC process, then a Gaussian distribution is used to create a background model and detect the distant moving objects, but this method only applicable to runway scene. Wang. J.M et al. [20] develops a 3D camera motion model, which can be applied to general case. In this paper, our electronic image stabilization is feature-based and a method for choosing the most accurate background feature points from a moving camera is proposed to stabilize the frames.

Various methods for detecting control points in an image have been developed, Schmid et al. [23] has surveyed and compared various point detectors, finding the Harris detector [24, 25]

to be most repeatable. Mikolajczyk [26] has proposed the scale-adapted Harris with automatic scale selection. However, this algorithm computes the image gradient based on discrete pixel differences, and finite differences can provide a very poor approximation to a derivative. In this work we resolve the above two shortcomings by applying optimal derivative filters. Optimal derivative filters method in general has emerged as an optimization of the rotation-invariance of the gradient operator [27]. It is aim to minimize the errors of in the estimated direction. We extend the application of optimal derivative filters to realize the accurate locations of the corners.

Feature matching is another important step in feature-based motion estimation. In recent years, several feature matching methods have successfully applied in image sequences motion estimation, such as invariant block matching [28] and feature point matching [29]. However, due to the noise and occlusion, some feature points displace even when their positions are detected with high accuracy. As a result, some matching points will be more accurate than others, which will affect the accuracy of video stabilization. The projective invariant [30] is a means to evaluate the geometrical invariability between images. This paper develops the projective invariant method that can distinguish more accurate matching points from the less accurate ones, and the most accurate points that belong to the planar background will be used to estimate transformation model.

The transformation model can be used to stabilize the video sequence by repositioning image frames in inverse direction of transformation model. However, digital image sequences acquired by airplane video camera are usually affected by unwanted positional fluctuations, which will affect the visual quality and impede the sub-sequent processes for various applications. Kalman filters [14] has been used to compensate the unwanted shaking of camera without the intentional camera motion. We will adopt a Sage-Husa Kalman filter [31] where the correction vector for each image frame is obtained as the difference between the filtered and original positions. This assumption helps to distinguish and preserve the intended camera motion.

2. Approach

When a video of a dynamic scene is captured by a moving aerial camera, knowing that two overlapping images are related by the projective transformation, a 2-D model can well trade off the accuracy and computational complexity for EIS. Assuming (x, y) represents a point in the base image and (X, Y) represents the same point in the image overlapping the base image, the projective transformation between the two points in the images can be written as:

$$\begin{cases} X = \dfrac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ Y = \dfrac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{cases}. \tag{1}$$

Having the coordinates of 4 corresponding points in the images, the unknown parameters $h_{11} - h_{33}$ of the transformation can be determined by substituting the corresponding points into (1) and solving the obtained system of the linear equations.

2.1 Feature Points Extraction Based on Optimal Derivative Filters

Feature detection is the first and critical step for image stabilization, and has been studied extensively in recent years [32]. We use Harris corner detector [24] in our stabilization framework because of the abundance of corners in aerial images. Harris detector, which uses the second-moment matrix as the basis of its corner detections, describes the curvature of the autocorrelation function in the neighborhood. For an image $I(x, y)$, the Harris detector based on the second-moment matrix can be expressed as:

$$M = \begin{bmatrix} G_{xt}^2 & G_{xt}G_{yt} \\ G_{yt}G_{xt} & G_{yt}^2 \end{bmatrix} * h, \tag{2}$$

where h is the Gaussian smoothing function. G is the traditional image gradient, which are given as follow:

$$\begin{bmatrix} G_{xt} \\ G_{yt} \end{bmatrix} = \begin{bmatrix} \partial I / \partial x \\ \partial I / \partial y \end{bmatrix} = \begin{bmatrix} I \otimes [d(k)]_{n=-L}^{L} \\ I \otimes \{[d(k)]_{n=-L}^{L}\}^{\mathrm{T}} \end{bmatrix}, \tag{3}$$

where $[d(k)]_{n=-L}^{L}$ is the general form of a linear phase Finite Impulse Response (FIR) and can be written as:

$$[d(k)]_{n=-L}^{L} = [d_L \cdots d_2 \; d_1 \; 0 \; -d_1 \; -d_2 \cdots -d_L]. \tag{4}$$

The Harris detector provides good repeatability under rotations and various illuminations; unfortunately, computing derivatives is sensitive to quantization noise, and the Harris corner detector has poor localization performance, particularly at certain junction types [32]. In this section, instead of using the traditional image gradients, we designed a new optimally first-order derivative filter with more accurate location and rotation-invariance. Optimal derivative filters method in general have emerged as an optimization of the rotation-invariance of the gradient operator [27]. It aims to minimize the errors in the estimated direction. We extended the application of optimal derivative filters to realize the accurate locations of the corners.

The Fourier transform of $d(n)$ is:

$$D(\omega) = \sum_{n=-L}^{L} d(n) \exp\{-jn\omega\} = 2j \sum_{n=1}^{L} d_n \sin(\omega n). \tag{5}$$

Ideally, our objective is to obtain the first-order derivative transfer function $D(\omega) = j\omega$. We can design the coefficients $\{d(n)\}_{n=1}^{L}$ to meet this function as closely as possible. Because the signal $I(x)$ and its derivative $I_x(x)$ are hard to get accurately in discrete domain, Here a pair of filters p and d is designed, and let $[I * p](x)$ and $[I * d](x)$ as an original and its derivative in a accurate form, respectively. We denote the filters pair by $P(\omega)$ and $D(\omega)$ in frequency domain. Then the error $j\omega P(\omega) - D(\omega)$ can be minimized by a more accurate method. Then the weighted least-squares error criterion for the designed filter is defined as:

$$E^2(p,d)=\frac{\int_{-\pi}^{\pi}\left[j\omega P(\omega)-D(\omega)\right]^2 d\omega}{\int_{-\pi}^{\pi}P^2(\omega)d\omega}. \tag{6}$$

This function is the form of Rayleigh quotient, so the result can be found using the Singular Value Decomposition(SVD). Then $p(n)*d(n)$ is called rotation-equivariant derivative filter [27], which shows good accuracy and rotation invariance. We choose 5-tap pair of filters to get good performance of precision. The resulting filter pair values are given in **Table. 1**.

Table 1. Matched pairs of prefilter p and derivative d kernels for a 5-tap

5-tap	p	0.037659	0.249153	0.426375	0.249153	0.037659
	d	0.109604	0.276691	0.000000	-0.276691	-0.109604

A standard image as show in **Fig. 2** (a) is distorted by rotations with angles ranging from $10°$ to $90°$, and by zooms with scales ranging from 0.6 to1.5, Assume that we choose a feature point (x_0, y_0) in the standard image, it is straightforward to know that the actual positions of feature point have changed to (x_1, y_1) and (x_2, y_2), respectively, in the rotated and scaled images. Knowing the correspondence point $(x_i^{'}, y_i^{'})$, we can get the Euclidean distance D between the point $(x_i^{'}, y_i^{'})$ and the actual point position (x_i, y_i). Simulation results are given in Fig.2 (d) and (e). It can be seen from the simulation results that the use of optimal derivative filters has consistently produced more accurate results than the use of Harris detector without optimal derivative filters.

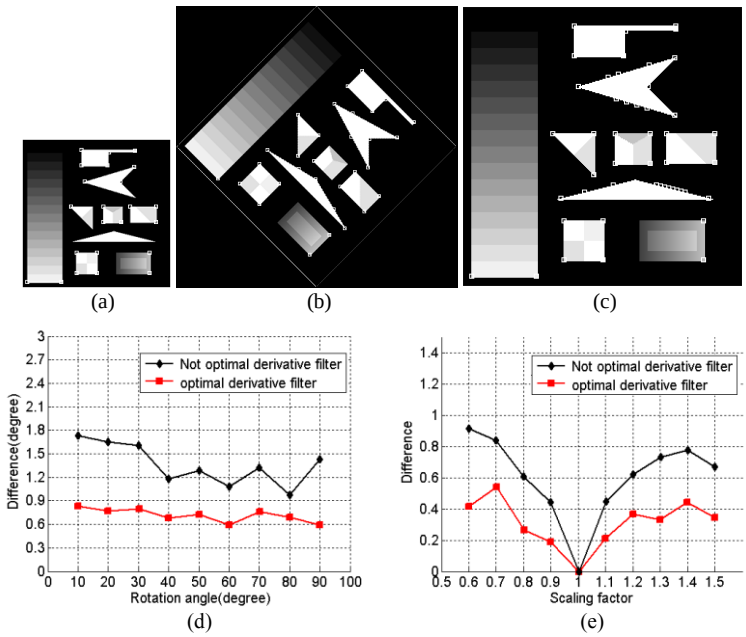


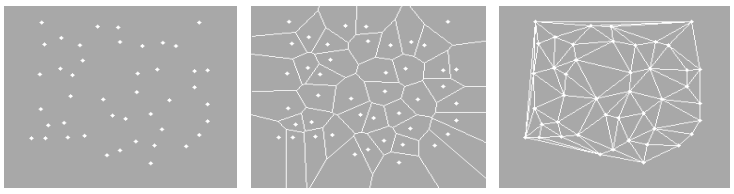
Fig. 2. Original image and the image after rotation and zooming. (a) Original image (126×126). (b) Image rotation $45°$. (c) Image zoom (250×250). (d) Corner error of the rotation angle. (e) Corner error of the scaling factor.

2.2 Correspondence between Points

After feature points have been detected, the next step is to find the correspondences between two point sets. The process involves removing the outliers and estimating the parameters of transformation. RANSAC algorithm is introduced by Fishler and Bolles in 1981 [33], and this algorithm uses a distance threshold to find the transformation matrix which maps the greatest number of point pairs between two images. Due to its ability to tolerate a large fraction of outliers, the algorithm is a popular choice for robust estimation of transformation matrix. Its lower-bound computation complexity is very low. However, it may not find the correspondences until after a large number of iterations computed, so the upper-bound computational cost of the RANSAC is substantial.

In a comparison study that involved several well known topological structures, the Delaunay Triangulation (DT) [34] was found to have the best structural stability under random positional perturbations.

For a set of points $p_1, p_2, \cdots p_n$, we obtain the DT by first calculating its Voronoi Diagram (VD). The VD of a set of points is the division of a plane or space into regions for each point. The regions contain the part of the plane or space which is closer to that point than any other. With a given VD, the DT is the straight line dual of the VD. A set of points are shown in **Fig. 3** (a), their VD is shown in Fig. 3 (b), while their DT is shown in **Fig. 3** (c).



(a) (b) (c)

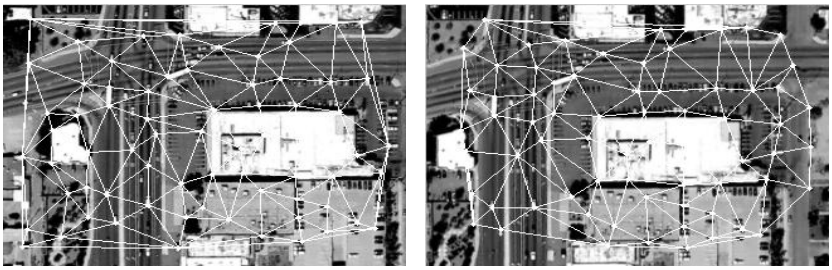
Fig. 3. (a) A set of points. (b) Voronoi Diagram. (c) Delaunay Triangulation.

In this research, we choose delaunay triangulation edges to find the matching pairs between feature points in overlapping images. An initial match between two point sets is obtained by selecting disjoint edge pairs in each DT that have the same transformation parameters. Before computing the parameters of the projective transformation, we will do some work to reduce the computation time. (i) For Delaunay Triangulation edges, the long DT edges are supposed to be more distinctive, and matching on long edges is considered to be more stable than matching the short ones. Therefore, we use only the longest 50 to 100 edges for feature matching. (ii)For aerial video images with general perspective changes (the viewpoints for the two images are not significantly different), the orientations of the corresponding edges in local areas should be relatively consistent. Therefore, we will use the edges with similar orientations (for example, the angle difference between the edge pairs is less than 10°). (iii) Strength contrast of corner response in a local region along both sides of the line can be used to further remove wrong candidates in the searching image. Assuming the equation expression of a line is $Ax + By + C = 0$, for a local region centered at the line, the average corner response on one side of the line is l_1, and the average corner response on the other side of the line is l_2. A strength contrast S for each line is assigned as $l_2 - l_1$, and then we have:

$$S = \begin{cases} 1 & l_2 - l_1 \geq 0 \\ -1 & otherwise \end{cases}. \tag{7}$$

If the strength contrast S of two matching edges are not equal, then the candidate edge is not considered as a possible matching edge and is excluded from further matching process.

An example of point matching in this method is given in **Fig. 4**. The Delaunay Triangulation edges obtained from the 100 stable corner points are shown in each image. We can see that many of the same DT edges are found in two images.



(a) (b)

Fig. 4. Two aerial frames show 100 stable corner points along with the Delaunay Triangulation.

2.3 Motion Parameters Estimation Based on Projective Invariant

The feature points detected by the optimized Harris detector are determined up to sub-pixel accuracy. Due to noise, 3-D structures or moving objects in image sequences, some feature points displace when their positions are detected by optimal derivative filters. As a result, there are certain matching points remain more invariant than others.

There exist some image properties that remain invariant under projective transformation. For projective transformation, the most fundamental invariant is called the cross-ratio invariant. The cross-ratio can be defined for four collinear points or five coplanar points, the five coplanar points is most suitable to our problem as we already have matching points in the image.

The five-point cross-ratio invariance is defined as follows [35]. Given five points $A_i, i = 1...5$ in an image, The cross-ratio of five points is defined as:

$$\lambda = \frac{(\Delta A_1 A_2 A_4)(\Delta A_1 A_3 A_5)}{(\Delta A_1 A_3 A_4)(\Delta A_1 A_2 A_5)}, \tag{8}$$

where $(\Delta A_1 A_2 A_4)$ is the oriented area of the triangle with vertices A_1, A_2 and A_4. Note that one point is shared by all four triangles and it is called the common point of the cross-ratio. It was shown [36] that the projective invariant of five points can be written as linear combination of four expressions:

$$\begin{aligned} J_1[\lambda] &= \frac{\lambda^6 - 3\lambda^5 + 3\lambda^4 - \lambda^3 + 3\lambda^2 - 3\lambda + 1}{\lambda^2(\lambda - 1)^2} \\ J_2[\lambda] &= \frac{2\lambda^6 - 6\lambda^5 + 9\lambda^4 - 8\lambda^3 + 9\lambda^2 - 6\lambda + 2}{\lambda^2(\lambda - 1)^2} \\ J_3[\lambda] &= 3, \quad J_4[\lambda] = -3 \end{aligned}. \tag{9}$$

The nontrivial projective invariant are unbounded function and can be written as:

$$J[\lambda] = \frac{J_2[\lambda]}{J_1[\lambda]} \quad . \tag{10}$$

If the feature point (x, y) in one frame and the coordinate point (X, Y) are related by the projective transformation, then by replacing (x_i, y_j) with (X_i, Y_j) in (8)-(10), we expect $J(x, y) = J(X, Y)$. If $J(x, y)$ and $J(X, Y)$ are not the same, the smaller their distance

$$D = \sqrt{[J(x, y) - J(X, Y)]^2} \tag{11}$$

Is, the higher the accuracy of the five matching points will be. Then we can select the best combination if the combination gives the smallest distance, and the best 5 matching points out of n will be selected to determine the parameters of the projective transformation.

An example using the projective constraint in image alignment is given in **Fig. 5**. As shown in Fig. 5 (a)-(b), the marked points on the local moving objects and on the 3-D structures are obviously inaccurate matching points, and these feature points will probably result in transformation model estimation error. The distance D in (11) is calculated for combination of 5 most accurate points that belong to the background, and the combination can produce the smallest distance. Fig. 5 (c) and (d) show absolute intensity difference of images registered using all the correspondences and using the best five correspondences, respectively. The difference between the two is significant.

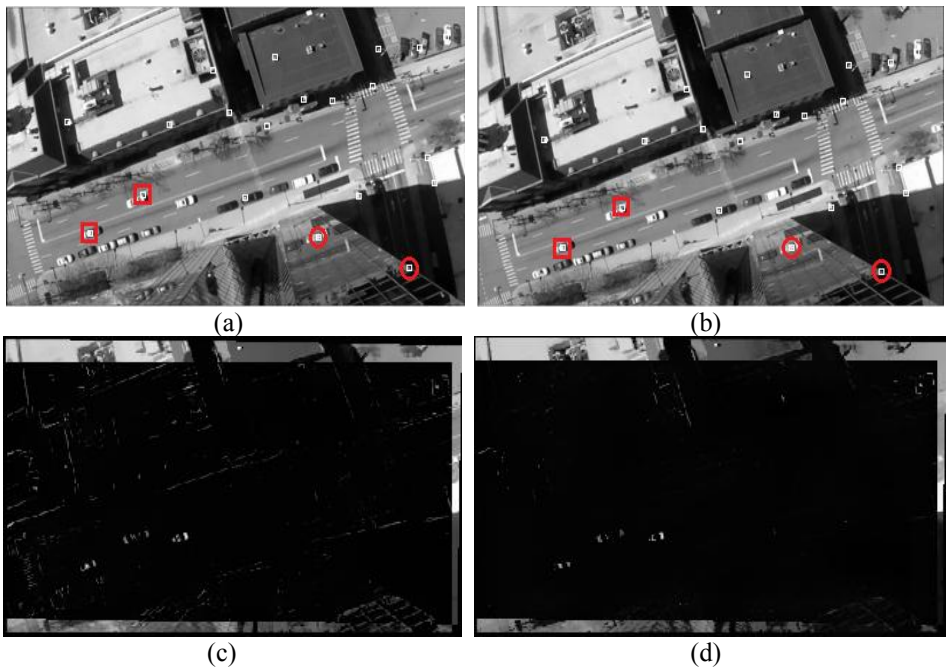


Fig. 5. (a), (b) Two images showing the matching feature points. (c) alignment result using all feature points. (d) alignment result using best five matching points.

2.4 Inter-frame Motion Compensation

When a video of a dynamic scene is captured by a moving camera, two types of motion will be

present in the video: one type is caused by the camera jitter and the second one is caused by the camera pan. Before motion compensation, it is clear that only the unwanted camera jitter should be removed by applying a low pass filter. Similar to Kalman filter [1], Sage-Husa filter [37] is based on the following assumption: the intentional camera scan is in a smooth motion in a fixed direction; by contrast, the unwanted camera jitter's variation and direction is more random. We can obtain the smooth motion component x_{filter} using adaptive filter, then the final jitter component x_{jilter} is the difference between original motion vector x_{raw} and smooth motion component, that is $x_{jilter} = x_{filter} - x_{raw}$.

Sage-Husa adaptive filter is designed on the basis of typical discrete Kalman filter. It takes advantage of measurement data to constantly modify system noise and measurement noise. The basic state estimate and update equations of Sage-Husa adaptive filter are given by:

$$\begin{cases} S(k|k-1) = F * S(k-1) \\ P(k|k-1) = F * P(k-1)F^T + \hat{Q}(k-1) \\ S(k|k) = s(k|k-1) + K(k) * \varepsilon(K) \\ \varepsilon(k) = Z(k) - H * S(k|k-1) \\ K(k) = P(k|k-1) * H^T (H * P(k|k-1) * H^T + \hat{R}(k))^{-1} \\ P(k|k-1) = (I - K(k) * P(k|(k-1))) \end{cases}, \tag{12}$$

where K represents the filter gain matrix; F is the state transfer matrix; H denotes the measurement matrix; R is the equivalent measurement noise matrix; Q refers to the equivalent state noise variance matrix; $P(k-1)$ is the prior state covariance matrix; $P(k|k-1)$ is the state predicting covariance matrix.

The estimating equations of $\hat{R}(k)$, $\hat{Q}(k)$ are given by:

$$\hat{R}(k) = (1-d_k)\hat{R}(k-1) + d_k(\varepsilon(k) * \varepsilon(k)^T - H * P(k) * H^T), \tag{13}$$

$$\hat{Q}(k) = (1-d(k))Q(k-1) + d(k)(K(k) * \varepsilon(k)^T K(k)^T + P(k|k) - F * P(k-1|k-1) * F^T), \tag{14}$$

where $d(k) = (1-b)/(1-b^k)$, b is the fading factor, and $0 < b < 1$.

Sage-Husa's adaptive Kalman filtering algorithm cannot estimate Q and R simultaneously when R and Q are all unknown. Possibility, measurement noise covariance matrix can easily cause filtering divergence phenomenon because of losing both positive definite form and semi-positive definite form, so the stability and convergence cannot be fully guaranteed.

In this article, the innovation sequence [31] is chosen to predict the residual error. Measurement of residual error is as follow:

$$\varepsilon(k)^T * \varepsilon(k) \le \gamma * Trace(H * P(k|k-1) * H^T + \hat{R}(k)), \tag{15}$$

where γ represents reserve coefficient. When $\gamma = 1$, filtering algorithm achieves the optimal estimation result:

$$\varepsilon(k)^T * \varepsilon(k) \le Trace(H * P(k|k-1) * H^T + \hat{R}(k)). \tag{16}$$

When the formula (16) is not satisfied, it's indicating that the actual error is γ times over the theoretical value. Here, the weighted coefficient $C(k)$ is considered to correct $P(k|k-1)$:

$$P(k|k-1) = C(k) * F * P(k-1) * F^T + \hat{Q}(k). \quad (17)$$

Substituting the formula (17) into (16), we obtain:

$$C(k) = \frac{\varepsilon(k)^T * \varepsilon(k) - Trace(H * Q(k) * H^T + R(k))}{Trace(H * F * P(k) * F^T * H^T)}. \quad (18)$$

From formula (12) to formula (18), we can get sage-husa adaptive Kalman filtering algorithm with innovation sequence.

3. Experimental Results

This section presents some examples and quantitative results of the proposed stabilization algorithm for aerial video sequences. The algorithm has been implemented in C++ and all experiments have been carried out on DELL Intel Xeon E5410 2.33-GHz desktop computer with 9GB of RAM, with Windows 7 Enterprise Professional Edition. **Fig. 6** shows 8 sets of the unmanned aerial vehicle (UAV) video sequences that come from the predator data of VSAM at Carnegie Mellon University (Fig.6,No.1-No.4) , DARPA video sequence (Fig.6,No.5) and our aerial video data (Fig.6,No.6-No.8) , with size 320×240, including rural roads, fields and urban buildings, Etc.

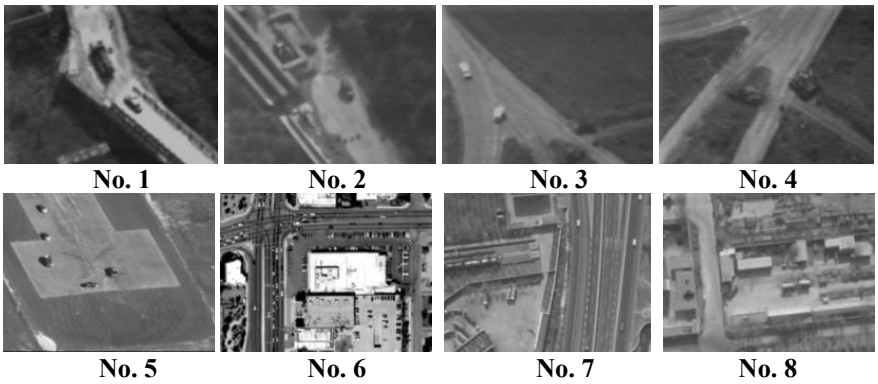
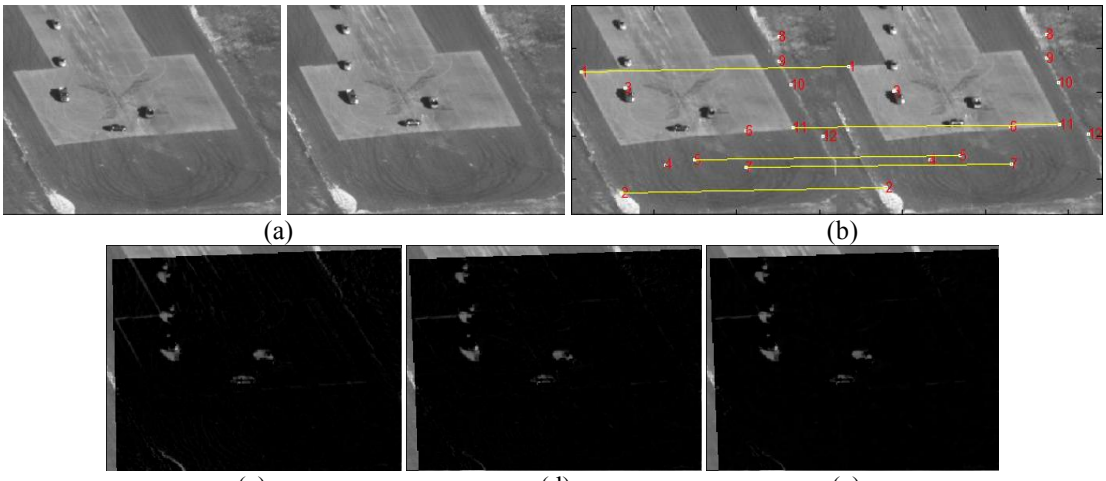


Fig. 6. UAV video sequences.

In order to demonstrate the accuracy of the motion estimation method that uses optimal derivative filters and projective invariant technique, we applied proposed method on two different types of video images: planar background and complex urban scene, which are shown in **Fig. 7** and **Fig. 8**.

An example using images of planar background is given in **Fig. 7**. The feature matching between the two frames is shown in **Fig. 7 (b)**. The total number of correspondences is 12. The identified correspondences are marked using the same numbers (drawn in red) in both images. The most accurate five correspondence points are shown by yellow lines. **Fig. 7. (c)** shows the absolute intensity differences of images using Harris detector and all correspondence points

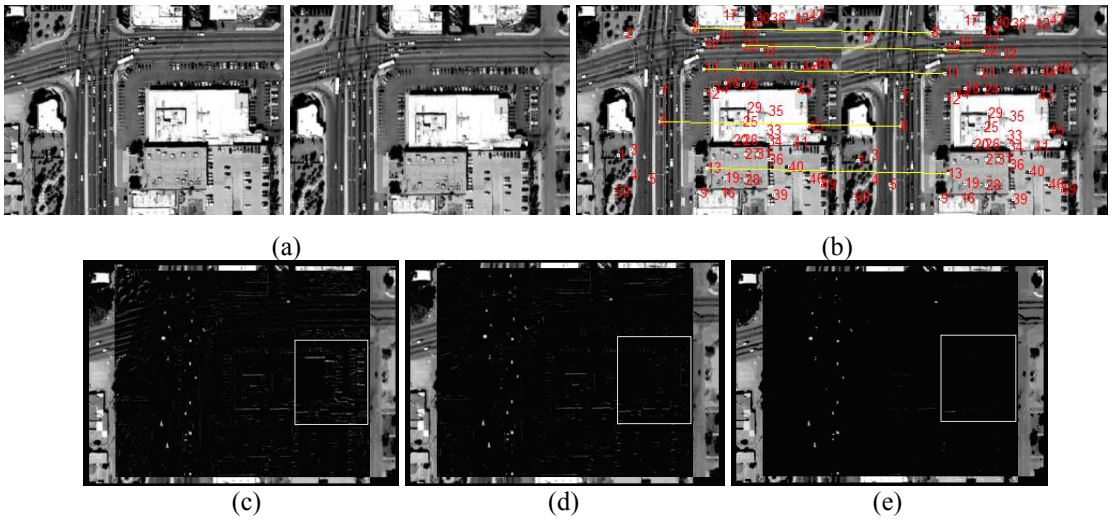
(HDAC), **Fig. 7 (d)** shows the alignment result using optimal derivative filters based Harris detector and all correspondence points (ODFAC), **Fig. 7 (e)** shows the alignment result using optimal derivative filters and best five correspondences obtained by projective invariant technique (ODFOI). Root-mean-squared (RMS) difference between aligned images when not using optimal derivative filters and projective invariant is 12.856. When using the optimal derivative filters method, the RMS difference between the images is 11.964, while RMS difference between images using optimal derivative filters and five matching points best satisfying the projective invariant is 11.018. We can find that high values show moving cars in all three alignment results, and the difference among the three results is small, because the scene is mainly composed of planar background and moving objects, but the result using optimal derivative filters and projective invariant is better than other two results.



(a) (b)

(c) (d) (e)

Fig. 7. Alignment of planar background. (a) Two aerial images. (b) The matching point pairs. (c)Alignment result using HDAC. (d) Alignment result using ODFAC. (e) Alignment result using ODFOI.



(a) (b)

(c) (d) (e)

Fig. 8. Alignment of complex urban scene. (a) Two aerial images. (b) The matching point pairs (c) Alignment result using HDAC. (d) Alignment result using ODFAC. (e) Alignment result using ODFOI.

The second example using images of complex urban scene is shown in **Fig. 8**. Most of the images have local distortion. The matching points are also shown in **Fig. 8 (b)**. The total number of correspondences was 40. The difference image after alignment using HDAC, ODFAC and ODFOI are shown in **Fig. 8 (c)**, **(d)** and **(e)**, respectively. The RMS differences when using HDAC, ODFAC and ODFOI are 15.031, 13.426 and 10.995, respectively. We can see that our algorithm produced more accurate transformation model parameters. Highest values show moving cars in **Fig. 8 (e)**. High values are also found at some rectangular areas and parked cars in **Fig. 8 (d)** and such errors can confuse the motion detection and tracking.

To examine the geometric fidelity of the motion estimation method, the cross-ratio invariance of four collinear points was used to determine the accuracy of motion estimation. The cross-ratio of four collinear points is the projective invariant of a quadruple of points. Given four collinear points p_1, p_2, p_3 and p_4 in one image, the cross-ratio is calculated by the following:

$$C_r = \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}}, \tag{19}$$

where Δ_{ij} denotes the Euclidean distance between two points p_i and p_j.

Supposing a line is drawn in the image, and four points are lying on the line. Firstly we calculate the cross-ratio C_r using the four collinear points, and then from three of the points in the aligned image, we calculate the location of the fourth point in the aligned image using cross-ratio invariance of four collinear points. Then, the distance between the calculated fourth point and the actual fourth point in the aligned image is used as error between the original image and the aligned image.

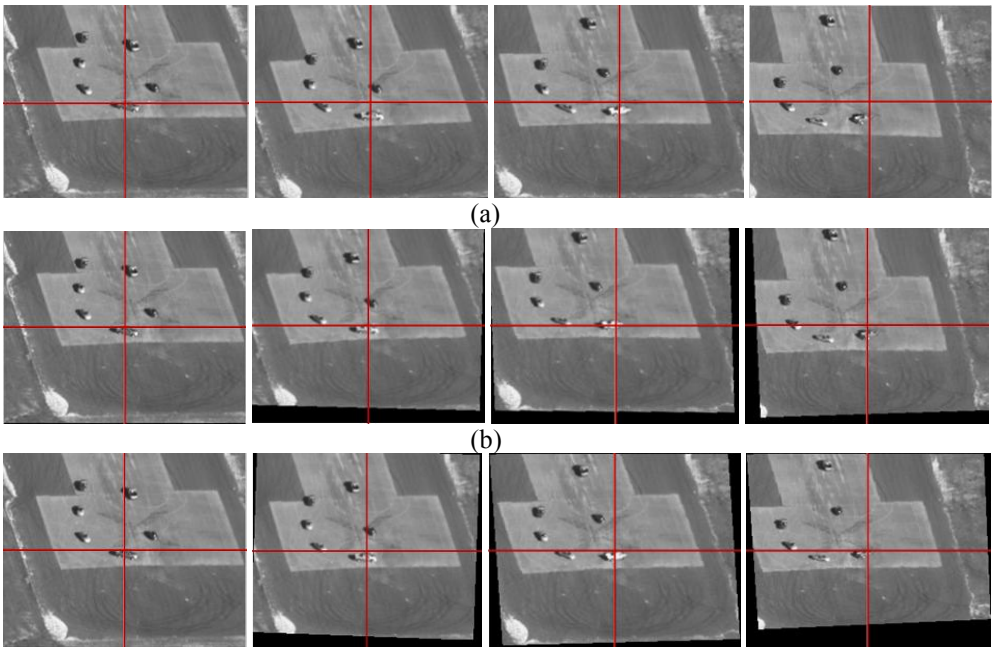
In order to evaluate the geometric fidelity of the motion estimation method, 8 sets of aerial video sequences were selected, as shown in Fig. 6, and each sequence contains 50 frames. **Table 2** shows more detailed test results of the feature points numbers and geometric fidelity errors using HDAC, ODFAC and ODFOI after registering frames. it is obvious that the average error using ODFOI is smaller than using HDAC and ODFAC, so we can obtain the more accurate alignment result.

Table 2. Comparison of the HDAC, ODFAC and ODFOI algorithms

	HDAC		ODFAC		ODFOI	
	Feature Point	**Average Error**	**Feature Point**	**Average Error**	**Feature Point**	**Average Error**
No.1	560	1.735	190	1.348	190	0.549
No.2	384	1.157	168	0.965	168	0.354
No.3	269	2.254	124	1.962	124	1.571
No.4	358	1.264	135	1.154	135	1.064
No.5	758	1.554	283	1.021	283	0.465
No.6	644	2.258	205	2.064	205	1.587
No.7	587	0.825	193	0.621	193	0.605
No.8	361	1.364	157	1.091	157	0.757
Average	490	1.551	181	1.278	181	0.869

To illustrate the stabilization results of the proposed algorithm, a comparison of video stabilization based on HDAC and ODFOI is depicted in **Fig. 9** and **Fig. 10**. Snapshot images of the planar background video sequence corresponding to Frames 30, 60, 90 and 120 are

illustrated in **Fig. 9 (a)**. Frames 90, 118, 160 and 189 of the complex urban scene video sequence are shown in **Fig. 10 (a)**. It's observed that the proposed algorithm well corrected the rotational and translational motions considering that there is intentional motion in the horizontal direction.

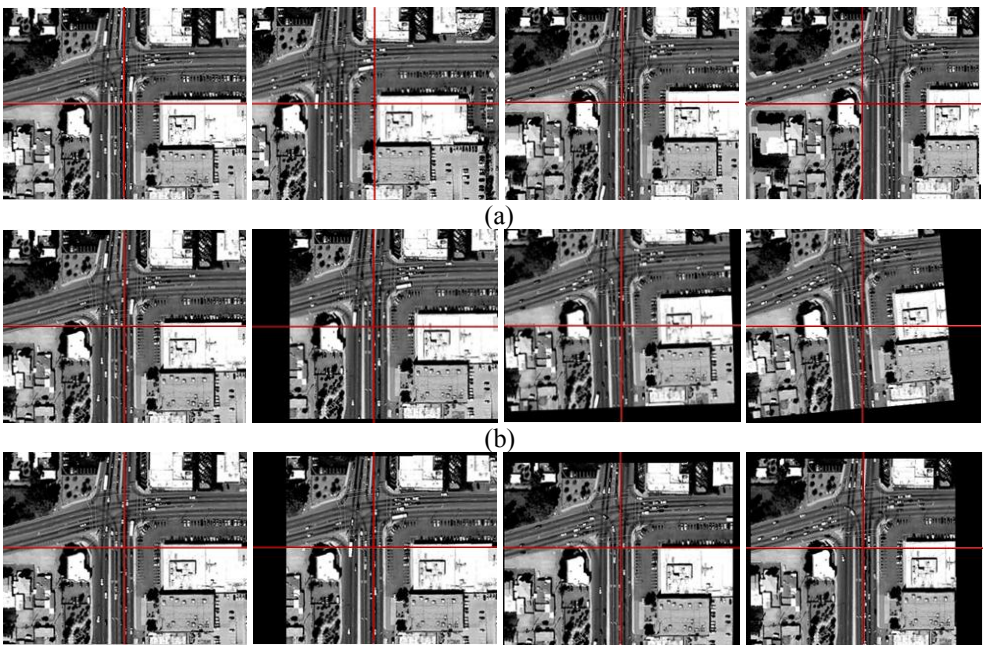


(a)

(b)

(c)

Fig. 9. Comparison of video stabilization for the planar background video sequence: (a) Original image. (b) Stabilization result using HDAC. (c) Stabilization result using ODFOI.



(a)

(b)

(c)

Fig. 10. Comparison of video stabilization for complex urban video sequence: (a) Original image. (b) Stabilization result using HDAC. (c) Stabilization result using ODFOI.

As discussed in section 2.4, intended aerial video is removed using Sage-Husa adaptive filter so that only unwanted jitter is removed during the stabilization process. **Fig. 11** shows how the method of estimating intended video motion presented in this paper performs on the UAV video of complex urban scene. The ODFOI method has a more steady change than HDAC method in the vertical direction, and successfully removes high-frequency jitter and smoothly follows the global motion trajectory.

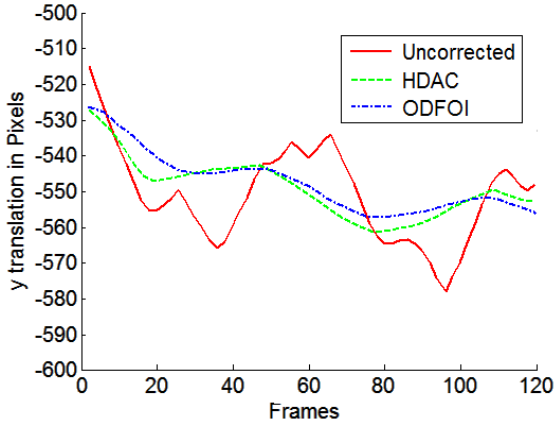


Fig. 11. The result of filtered y translation motion vectors

To make an objective evaluation of the image stabilization method between the stabilized image and the reference image, the peak signal-to-noise ratio (PSNR) can be used as a measure. The larger the value of PSNR is, the smaller the inter-frame error is. The PSNR between consecutive images $(M \times N)$ It and It+1, called global transformation fidelity (GTF) , is defined as

$$PSNR(I_t, I_{t+1}) = 10 \lg \frac{255^2}{MSE(I_t, I_{t+1})}, \tag{20}$$

$$MSE = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} (I_t(m,n) - I_{t+1}(m,n))^2, \tag{21}$$

where M and N are the width and height of the images, respectively; MSE denotes the mean square error calculated for the considered images.

The GTF index was used to evaluate motion compensation with respect to an initial reference image. **Fig. 12** displays the PSNR curves of **Fig. 6**. no 6 for considered system. Each point of the curves was calculated by varying the motion range of the image to be stabilized. The lower PSNR curve and upper PSNR curve are the GTF of the original and stabilized video sequences, respectively. As can be seen from the GTF, the curve that represents the uncompensated sequence is always below the compensated line. This means that the proposed method is able to compensate for unwanted motion. The PSNR values for the planar background video sequence and complex urban video sequence are listed in **Table. 3,** which are computed over 100 frames. It is observed that the PSNR of the ODFOI method is smaller than that of the HDAC, which means that the proposed method more robust to irregular conditions than that of the HDAC.

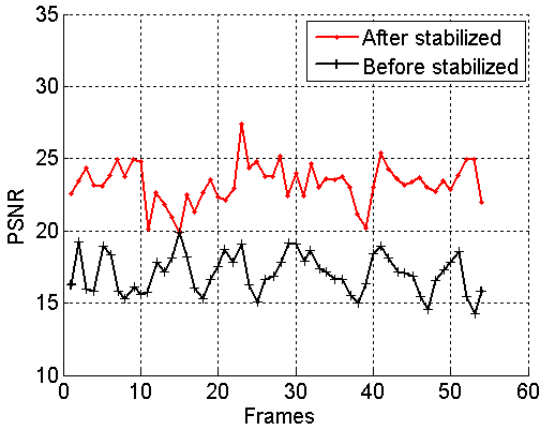


Fig. 12. Comparison of inter-frame PSNR curve

Table 3. PSNR of test video sequence

Video Seq.	ODFOI	HDAC
planar background video	28.372	23.145
complex urban video	23.864	19.691

The computation complexity of the proposed stabilization method is a funciton of image size, the number of video frames, the number of feautre points detected in each iamge, and the nmber of obtained correspondences. For the same number of feature points, the larger the image size, the more computation time will be needed to calculate the corners. Given an image of size $M \times N$ pixels, the computational complexity of the corner detector is on the order of $O(MN)$. If A and B feature points are obtained in two frames, the computational complexity of the matching algorithm to find the correspondences is on the order of $O(A^2B^2)$. If p correspondences are found, the computational complexity of the projective constraint that finds the best 5 correspondences is on the order of $O(p^5)$. The best 5 correspondences are then used to calculate the projective transformation parameters to register the images.

The images used in this study had $M = 320$ rows and $N = 240$ columns. 100 feature points were selected in each image. These parameters produced about a dozen correspondences by the matching algorithm. The processing time of the proposed method is less than 50 ms per frame based on a 3.2 GHz computer.

4. Conclusion

A new alignment method for stabilizing video frames captured by a moving aerial camera was described. The effectiveness of our approach has been demonstrated through a series of experiments in critical conditions and the experimental results show that the proposed scheme carries out real-time aerial video stabilization under complex environments with change of scenes, and achieves precision stabilization. Future work will be devoted to extend the proposed method to stabilize very large dynamic scenes with non-planar background, for example, apply the algorithm to stabilize sub-images within the frames. Moreover, further investigations will incorporate proposed method into various applications such as textured image classification and moving objects tracking.

References

[1] Andrey Litvin, Janusz Konrad and William C. Karl, "Probabilistic video stabilization using Kalman filtering and mosaicking," *IS&T/SPIE Symposium on Electronic Imaging, Image and Video Communications and Proc*, 2003. Article (CrossRef Link)

[2] Yasuyuki Matsushita, Eyal Ofek, Xiaoou Tang and Heung-Yeung Shum, "Full-frame Video Stabilization," *Microsoft Research Asia. CVPR,* 2005. Article (CrossRef Link)

[3] Feng Liu and Hailin Jin, "Content-preserving warps for 3D video stabilization," *Proceeding SIGGRAPH'09 ACM SIGGRAPH*, vol.28, no.3, 2009. Article (CrossRef Link)

[4] Lucio Mercenaro, Gianni Vernazza and Carlo S. Regazzoni, "Image stabilization algorithms for video-surveillance applicaion," *In Proc. ICIP*, pp.349-352, 2001. Article (CrossRef Link)

[5] Tao Zhao and Ram Nevatia, "Car detection in low resolution aerial images," *Image and vision computing*, vol.21, no.8, pp.693-703, 2003. Article (CrossRef Link)

[6] Gary. F.Templeton, "Video image stabilization and registration technology," *communications of the ACM*, vol.49, no.2, pp.15-18, 2006. Article (CrossRef Link)

[7] Chi-Han Chuang,Yung-Chi-Lo and Chin-Chun Chang, "Multiple object motion detection for robust image stabilization using blocak based hough transform," *IIH-MSP*, pp.623-625, 2010. Article (CrossRef Link)

[8] Sheng-Che Hsu, Sheng-Fu Liang and Chin-Teng Lin, "A robust digital image stabilization technique based on inverse triangle method and background detection," *Transactions on Consumer Electronics*, vol.51, no.2, pp.335-345, 2005. Article (CrossRef Link)

[9] Puglisi Giovanni and Battiato Sebastiano, "A robust image alignment algorithm for video stabilization purposes," *Transactions on circuits and systems for video technology*, vol.21, no.10, pp.1390-1401, 2011. Article (CrossRef Link)

[10] Battiato Sebastiano and Rastislav Lukac, "Video stabilization techniques," *Encyclopedia of Multimedia. New York: Springer-Verlag*, pp. 941–945, 2008. Article (CrossRef Link)

[11] Ben Tordoff and David W Murray, "Guided sampling and consensus for motion estimation," *European Conference n Computer Vision*, 2002. Article (CrossRef Link)

[12] Filippo Vella, Alfio Castoorina, Massimo Mancuso and Giuseppe Messina, "Digital image stabilization by adaptive block motion vectors filtering," *IEEE Trans. on Consumer Electronics*, vol.48, no.3, pp. 796–801, 2002. Article (CrossRef Link)

[13] Sung-Jea Ko, Sung-Hee Lee and Kyung-Hoon Lee, "Digital image stabilizing algorithms based on bit-plane matching," *IEEE Trans. on Consumer Electronics*, vol.44, no.3, pp. 617–622, 1998. Article (CrossRef Link)

[14] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation," *IEEE Trans. on Consumer Electronics*, vol. 49, no.4, pp.1320–1325, 2003. Article (CrossRef Link)

[15] Chen. H., Liang. C.K., Peng. Y.C., Chang. H.A, "Integration of digital stabilizer with video codec for digital video cameras," *Trans.Circuits Syst. Video Technol*, vol.17, no.7, pp. 801-813, 2007. Article (CrossRef Link)

[16] Sebastiano Battiato, Arcangelo Ranieri Bruna and Giovanni Puglisi, "A robust block based image/video registraiton approach for mobile imaging devices," *Trans.Multimedia*. vol.12, no.7, pp.622-635, 2010. Article (CrossRef Link)

[17] Juan-Juan Zhu and Bao-Long Guo, "Global point tracking based panoramic image stabilization system," *Optoelectronics Letters*, vol.5, no.1, pp.61-63, 2009. Article (CrossRef Link)

[18] Junlan Yang, Dan Schonfelda and Magdi Mohamed, "Robust video stabilization based on particle filter tracking of projected camera motion," *Trans.Circuits Syst.Video Technol*, vol.19, no.7, pp.945-954, 2009. Article (CrossRef Link)

[19] Cheng-Hua Pai, Yu-Ping Lin and Gerard G. Medioni, "Moving Object Detection on a Runway Prior to Landing Using an Onboard Infrared Camera," CVPR'07, pp.17-22, 2007. Article (CrossRef Link)

[20] J.M.Wang, H.P.Chou, S.W. Chen and C.S. Fuh, "Video stabilization for a hand-held camera Based on 3D Motion Model," *ICIP*, pp.7-10, 2009. Article (CrossRef Link)

[21] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol.60, no.2, pp.91-110, 2004. Article (CrossRef Link)
[22] Lianming Xu, Zhongliang Deng and Ling Fang. "Research on an Anti-Perturbation Kalman Filter Algorithm," *Journal of Networks*, vol.6, no.10, pp.1430-1436, 2011. Article (CrossRef Link)
[23] Cordelia Schmid, Roger Mohr and Christian Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis*., vol.37, no.2, pp.151-172, 2000. Article (CrossRef Link)
[24] Chris Harris and Mike Stephens. "A combined corner and edge detector," *Alvey vision conference*, pp. 147-152, 1988. Article (CrossRef Link)
[25] Qing Zhu, Bo Wu, and Neng Wan, "A Subpixel location method for interest points by means of the harris interest strength," *Photogrammetric Record*, vol.22, no.120, pp. 321-335, 2007. Article (CrossRef Link)
[26] Krystian Mikolajczyk and Cordelia Schmid, "An affine invariant interest point detector," *in Proc. of European Conference on Computer Vision, ECCV*, pp.128-142, 2002. Article (CrossRef Link)
[27] Hany Farid and Eero. P. Simoncelli, Differentiation of discrete multi-dimensional signals, IEEE Transctions on Image Processing, vol.13, no.4, pp. 496-508, 2004. Article (CrossRef Link)
[28] Lidong Xu and Xinggang Lin, "Digital image stabilization based on circular block matching," *IEEE Trans. on Consumer Electronics*, vol.52, no.2, pp.566-574, 2006. Article (CrossRef Link)
[29] Qing Zhu, Yunsheng Zhang, Bo Wu and Yeting Zhang, "Multiple Close-range Image Matching Based on a Self-adaptive Triangle Constraint," *Photogrammetric Record*, vol.25, no.132, pp.437-453, 2010. Article (CrossRef Link)
[30] Emanuele Trucco, "Geometric Invariance in Computer Vision. Cambridge," *MA: MIT Press*, 1992. Article (CrossRef Link)
[31] Lianming Xu, Zhongliang Deng and Ling Fang, Research on an Anti-Perturbation Kalman Filter Algorithm," *Journal of Networks*, vol.6, no.10, pp.1430-1436, 2011. Article (CrossRef Link)
[32] Tinne Tuytelaars and Krystian Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol.3, no.3, pp.177-280, 2008. Article (CrossRef Link)
[33] Martin A. Fischler and Robert C. Bolles, "Random sample consensus:a paradigm for model fitting with applications to image analysis and automated cartography," *Commun.ACM*, pp.381-395, 1981. Article (CrossRef Link)
[34] Franco P. Preparata and Michael I. Shamos, "Computational geometry: An introduction," *New York: Springer-Verlag*, pp.95-226, 1985. Article (CrossRef Link)
[35] Peter Meer, Sudhir Ramakrishna and Reiner Lenz, "Correspondence of coplanar features through P2-invariant representations," *Applications of Invariance in Computer Vision*, vol.825, pp. 473-492, 1994. Article (CrossRef Link)
[36] Chun-ming Xie, Yan Zhao and Ji-nan Wang, "Application of an improved adaptive Kalman filter to transfer alignment of airborne missile INS," *Proc. SPIE*, vol.7, no.129, pp.260-266, 2008. Article (CrossRef Link)

Meng yi (S'12) received the M.S. degree in Electrical Engineering from Northwestern Polytechnical University, Xi'an, China, in March 2008. Since 2009, he has been a Ph.D. of Electric circuit and systematic at Xidian University. Currently, he is a visiting doctoral candidate in Department of Electrical and Computer Engineering and Center for Automation Research, University of Maryland, College Park, maryland, USA. His research interests include computer vision, pattern recognition, signal processing and biometrics.

Baolong Guo received the M.S. and Ph.D. degrees from Xidian University in 1988 and 1995, respectively, all in communication and electronic system. From 1998 to 1999, he was a visiting scientist at Doshisha University, Japan. He is currently a full professor with the Institure of Intelligent Control and Image Engineering(ICIE) at Xidian University.His research interests include neural networks, pattern recognition, and image processing.