2017 International Conference on Vision, Image and Signal Processing

Panorama Stitching, Moving Object Detection and Tracking in UAV Videos

Quanlu Wei,Songyang Lao and Liang Bai
Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China
e-mail: blessyou668@163.com, laosongyang@vip.sina.com, xabpz@163.com

***Abstract*—Unmanned Aerial Vehicles(UAV) are more and more wildly used recent years in many fields. It's convenient to acquire more static and dynamic information by uav aerial videos to grasp the scene situation. Frames registration, panoramic image mosaic, moving objects detection and tracking are the key and foundation of the aerial video analysis and processing. Firstly, we use a l_q-estimation method to remove the outliers and match the feature points robustly. Then we utilize a Moving Direct Linear Transformation (MDLT) method to find the homography of the frames more accurately, and stitch the frame sequence to a panorama. Finally, we apply a 5-frame difference method on the warped frames to detect the moving objects, and use a long-term visual tracking method to track the object of interest in complex scenes. The experiments show that our method achieve good results in different conditions.**

Keywords-registration; stitching; moving object detection; l_q-estimation; MDLT; visual tracking.

I. INTRODUCTION

Compared to the manned aircraft, the UAV are smaller, lighter, cheaper, and more suited to execute danger tasks. Small UAV equipped with visual sensor is an ideal platform for anti-terrorism, traffic monitoring, disaster relief, battlefield surveillance and so on [1,2]. Panoramic image mosaic, moving objects detection and tracking are the key technologies to completing these missions. Due to the movement of the platform, besides the foreground objects, the background is also moving, so the motion compensation of the background is a necessary step of UAV aerial video analysis and processing [3]. After the reception of the background motion compensation module registers the video frames and generates corresponding aligned images [4], the frames sequence can be stitched together to generate a panorama to grasp the overall information [5,6]. In addition, the moving objects can be detected by frame subtraction, and the object of interest is tracked by a tracking module. In this paper, we first extract the Harris features [7] of the adjacent frames, then introduce a robust method based on l_q -estimator for outliers removing and robust feature matching [8], after that, utilize a MDLT method to find the homography of the frames [9]. According to the homography, the frames are registered and the ego-motion of the platform is compensated, also the panorama is stitched by the frames sequence. Finally, the moving objects detection is accomplished by using a 5-frame subtraction method [10], and the object of interest is tracked by a visual tracking algorithm based on correlation filters [11].

II. IMAGE REGISTRATION

Image registration is to find the right positions of the corresponding feature points in two images by using a matching strategy, and then obtain the homography between two images for registration. In this paper, the image registration algorithm includes several parts: the extraction and description of Harris feature points, l_q -estimator for robust feature matching, MDLT method for estimating the homography. First, the feature points are extracted and described from the two image frames, and the matched points are obtained, then the outliers are removed and the features are matched robustly by l_q -estimator. Finally, we use MDLT method to weighted estimate the homography which satisfied the different parts of the image, and get the accurate projection model parameters to register the image frames.

A. l_q-Estimation for Feature Matching

For an image pair to be matched, we perform a feature matching method such as Harris to determine N initial matching correspondences

$$\{(x_n,y_n)\}_{n=1}^{N} \tag{1}$$

Where x_n and y_n are the 2-D coordinates of the matched feature points, (x_n,y_n) satisfies the following relationship if it is an inlier:

$$y_n = T(x_n) \tag{2}$$

The transformation $T(\cdot)$ can be estimated by least squares method with the K inlier matches [8]

$$\arg\min_{T} \sum_{n=1}^{K} (y_n - T(x_n))^2 \tag{3}$$

But there may be outliers in these points, the outliers should be removed to estimate the transformation correctly. Current methods usually use a two-step strategy or a hypothesize-and -verify technique such as RANSAC to solve the problem, these methods is always very time-consuming even can't get a reasonable result.

The robust feature matching method based on

978-1-5386-0612-4/17 $31.00 © 2017 IEEE
DOI 10.1109/ICVISP.2017.13

l_q-estimator directly estimates the transformation from initial correspondences with outliers. In order to classify the residual vector into an outlier set and an inlier set automatically, the classical least-squares cost is sensitive to outliers. l_0-norm is suitable for solving such a problem, but it is unreliable due to the noise contained in the observations. Usually the l_1-norm is adapted as the closest convex relaxation of l_0-norm to make a tradeoff. The l_q-estimator is more robust and effective for feature matching. The cost function is

$$\arg\min_{T} \sum_{n=1}^{N} \|y_n - T(x_n)\|_q^q \quad (0 < q < 1) \qquad (4)$$

where $\|\cdot\|_q$is a l_q-norm operator.

The outliers will be removed by applying the global transformation to the initial feature points.

B. Homography Estimation

For the low altitude aerial videos, the views between the frames do not differ purely by rotation or are not of a planner scene, using a basic homographic warp inevitably yields misalignment or parallax errors. The APAP(As-Projective-As-Possible) image stitching method which proposed by Julio Zaragoza etc. assumed that the details of the image satisfy different homography, and used a location dependent homography to warp each pixel, weighting estimated the homography by using MDLT method [9], which can alleviate the effect of misalignment and parallax errors.

Direct Linear Transformation(DLT) is a basic method to estimate the homography from a set of noisy point matches. Only two of the rows are linearly independent after vectorizing the homograph matrix into a vector, let a_i be the first-two rows of the LHS matrix computed for the i-th point match. Stacking vertically a_i for all i into matrix **A**[9].

Then the optimization objective is

$$\hat{\mathrm{h}} = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} \|a_i h\|^2 = \underset{h}{\operatorname{argmin}} \|Ah\|^2 \quad s.t.\ \|h\| = 1 \qquad (5)$$

The whole image just uses one homography reconstructed from $\hat{\mathrm{h}}$ for warping.

The MDLT method improved by estimating the homography from the weighted problem,

$$h_* = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{N} \left\|\omega_*^i a_i h\right\|^2 = \underset{h}{\operatorname{argmin}} \|W_* A h\|^2 \quad s.t.\ \|h\| = 1 \qquad (6)$$

The weights ω_*^i give higher importance to i-th point match that are closer to x_*.

$$\omega_*^i = \exp(-\|x_* - x_i\|^2/\sigma^2) \qquad (7)$$

$$W_* = \operatorname{diag}([\omega_*^1\ \omega_*^1\ \omega_*^2\ \omega_*^2 \cdots \omega_*^N\ \omega_*^N]) \qquad (8)$$

To prevent numerical issues in the estimation, they offset the weights with a small value γ within 0 and 1.

$$\omega_*^i = \max(\exp(-\|x_* - x_i\|^2/\sigma^2), \gamma) \quad \gamma \in [0,1] \qquad (9)$$

Calculating the homography of each pixel is unnecessarily wasteful. We thus uniformly partition the image into a grid of several cells, and take the center of each cell as x_*[9].

III. PANORAMA STITCHING

After the panoramic stitching of the aerial video frames, we can get the static image of large-scale scene to grasp the overall information.

Firstly, we warp the two frames to be stitched using the homography, map the pixels to the location in the panorama, stitch the frames successively, and then, fuse the two warped images to avoid the obverse difference near the seam-line. Commonly, the overlap rate of the adjacent frames, in the practical applications, we select the frames of a certain time interval according to the moving speed for stitching, which can reduce the computational complex. For the earlier stitched image, we only select the last frame not the whole stitched image to extract the feature points also for the computation speed.

IV. MOVING OBJECT DETECTION

For the aligned frames, an improved 5-frame difference method is used to detect the moving object. The traditional 3-frame difference method can detect the basic contour of the object, but the contour is always discontinuous, and the overlap of the objects is not easy to detect. According to the theory of the frames difference method, information fusion by multi-frame difference can be useful for extracting a more complete moving object. 5-frame difference method can partly overcome the deficiency of the 3-frame difference method[10]. For the 5 adjacent frames $f_{t-2}(x,y)$, $f_{t-1}(x,y)$, $f_t(x,y)$, $f_{t+1}(x,y)$, $f_{t+2}(x,y)$, we firstly use the median filter to remove salt and pepper noise, then do differential operation between the middle frame and the other 4 frames respectively. The result is as follows:

$$D_{13}(x,y) = |f_t(x,y) - f_{t-2}(x,y)| \qquad (10)$$

$$D_{23}(x,y) = |f_t(x,y) - f_{t-1}(x,y)| \qquad (11)$$

$$D_{34}(x,y) = |f_t(x,y) - f_{t+1}(x,y)| \qquad (12)$$

$$D_{35}(x,y) = |f_t(x,y) - f_{t+2}(x,y)| \qquad (13)$$

After filter the differential result, we introduce Otsu dynamic threshold segmentation method to obtain the binary images, and then use the "and" operation to restrain objects overlapping problem.

$$D_1 = D_{13}(x,y) \cdot D_{35}(x,y) \qquad (14)$$

$$D_2 = D_{23}(x,y) \cdot D_{34}(x,y) \qquad (15)$$

Then we use "or" operation on D_1 and D_2 to avoid bringing holes in the object contour.

$$\mathrm{D} = D_1 + D_2 \qquad (16)$$

The binary images may also have noise and small holes, which can get wrong bounding box of the objects. Finally, the moving object regions can be masked by morphological eroding and dilating to remove the noise and fill the holes, then the positions and scale of the objects can be obtained.

V. OBJECT TRACKING

There will be several object regions detected by the moving object detection step, we only choose one target of interest, use the long-term visual tracking algorithm based on correlation filters [11] to track the chosen object to obtain the position and scale of the object in real time. The tracker is initialized by the bounding-box detected in the detection operation.

The long-term visual tracking algorithm based on correlation filters is under the framework of kernelized correlation filter tracker, integrates Histogram of Oriented

Gradient, Color Naming and intensity to create the robust object appearance model. In the subsequent frames, the new position and scale of the object can be estimated by maximizing the correlation score of the translation filter and the scale filter respectively, and the filters are updated by the new position and scale. Meanwhile, we detect the tracking status in real time, and uses an online CUR filter to re-detect the object in case of tracking failure. The algorithm is robust to complex scenes for long term visual tracking. The flowchart of tracking is shown in Figure 1.

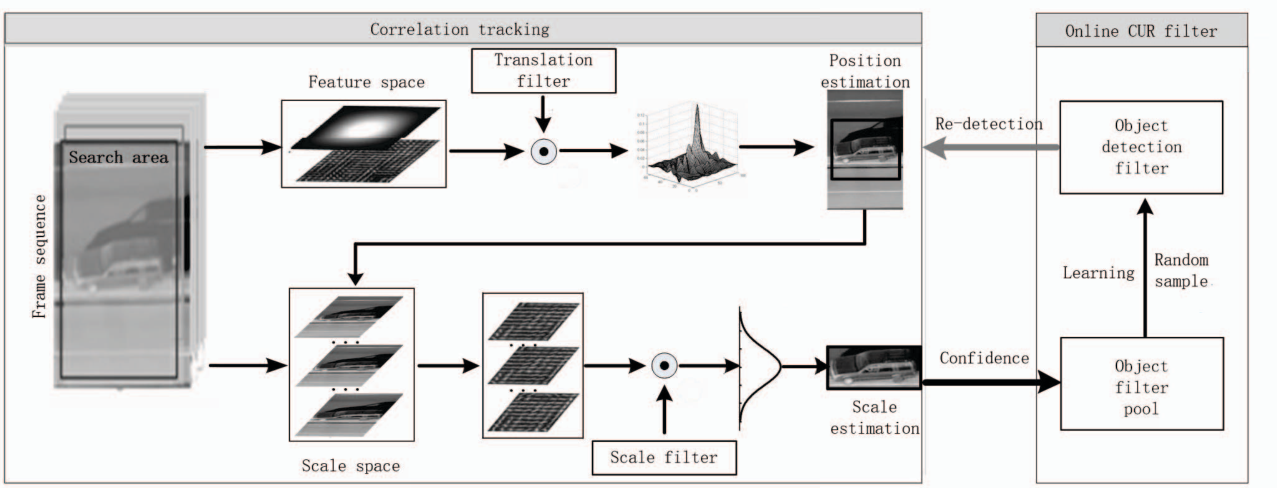


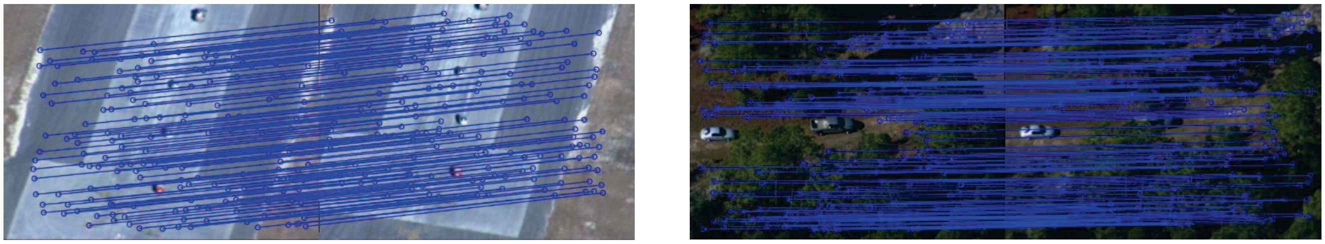
Figure 1. The flowchart of the tracking algorithm

VI. Experiments

The datasets in the experiments are selected from the aerial video data in the Video Verification of Identity(VIVID) public datasets proposed by DARPA. These datasets include background of less-textured and well-textured videos. We select two typical videos *egtest01* and *egtest05* for our experiments. The resolution ration is 640x480, frame rate is 30fps. The experiments are implemented in MATLAB R2016a on an Intel Core i5-7300HQ, 2.5GHz CPU, 8GB RAM computer.

A. Result of Image Registration

We first extract the Harris features in the 2 frames to be registered, use the Euclidean distance of the descriptors for rough matching, and then use the l_q-estimation method to remove the outliers. The final matching results are showed in Figure 2.



(a) Matching result of *egtest01*

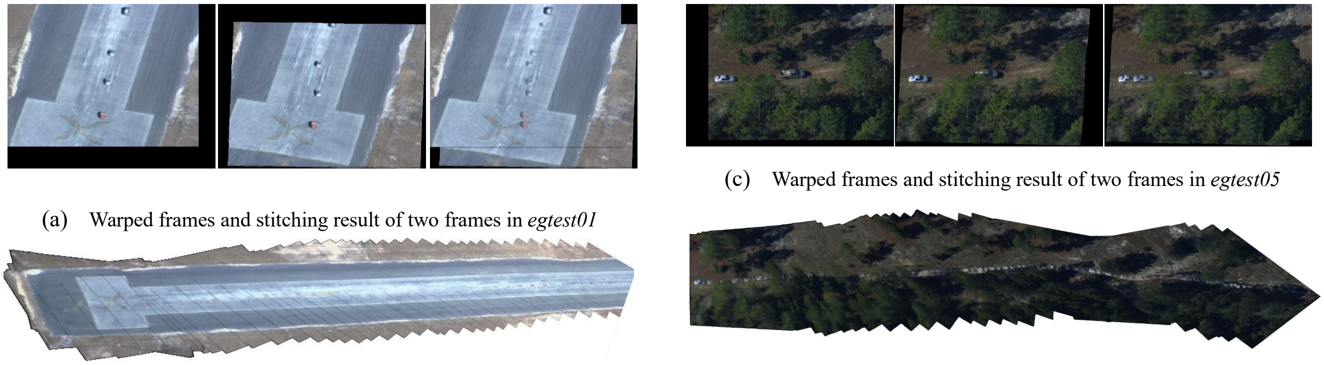
(b) Matching result of *egtest05*

Figure 2. Matching results of the feature points

We can see from figure 2 that the feature points are distributed uniform whether in less-textured or well-textured background scenes, and the points are mainly concentrated in the background, which will help to get the accurate registration results.

B. Result of Panoramic Stitching

We use MDLT method to obtain the transformation of 2 frames according to the matched feature points, then warp the image, stitch and fuse the reference image and the warped image. Finally, the frames are stitched together successively to get a panorama showed in Figure 3.



(a) Warped frames and stitching result of two frames in *egtest01*

(b) Panorama of *egtest01*

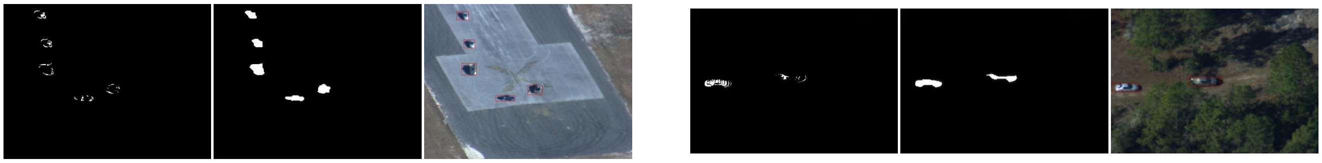
(c) Warped frames and stitching result of two frames in *egtest05*

(d) Panorama of *egtest05*

Figure 3. The results of warping and stitching

C. Result of Moving Object Detection

After the registration of the frames, we use 5-frame difference method to get the difference result, then remove the thin square noise, do morphological operation to get the object regions, finally the positions and the scales of the moving objects can be obtained from the bounding-boxes of the regions, which are showed in Figure 4.



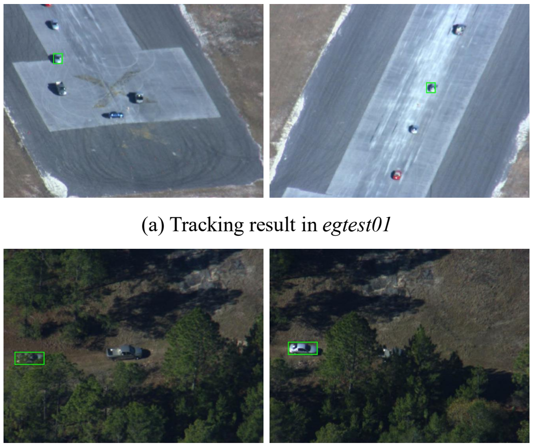
(a) Results of difference, morphological operation, and bounding-box in *egtest01*

(b) Results of difference, morphological operation, and bounding-box in *egtest05*

Figure 4. The results of moving objects detection

D. Result of Tracking

The tracking algorithm can update the scale of the object and re-detect the lost object in case of complete occlusion or out-of-view. Figure 5 shows the tracking results in different frames.



(a) Tracking result in *egtest01*

(b) Tracking result in *egtest05*

Figure 5. Results of tracking

VII. Conclusions

We have developed the implementations of a number of key image processing algorithms for aerial reconnaissance based on small UAV platform. The algorithms include registering the video frames, using the frames difference for moving objects detection, stitching the frames to panorama, tracking one of the detected objects. The experiment results show that the proposed method can work well in both less-textured background and well-textured background scenes for registration, stitching, detection, and complex scenes for tracking.

REFERENCES

[1] N. Heinze, M. Esswein, W. Kruger, G. Saur, "Automatic image exploitation system for small UAVs". Airborne Intelligence, Surveillance, Reconnaissance(ISR) Systems and Applications V, Proc. of SPIE, Vol. 6946,69460G ,2008

[2] Saad Ali, Mubarak Shah, "COCOA: Tracking in Aerial Imagery", Airborne Intelligence, Surveillance, Reconnaissance(ISR) Systems and Applications III, Vol.6209,62090D,2006

[3] Ahlem Walha, Ali Wali, Adel M. Alimi, “Video stabilization with moving object detecting and tracking for aerial video surveillance”, Multimedia Tools and Applications 74(17):1-23,2014

[4] Binpin Su, Honglun Wang, Xiao Liang and Hongxia Ji, “Moving Objects Detecting and Tracking for Unmanned Aerial Vehicle”, Foundations and Practical Applications of Cognitive Systems and Information Processing, 317-333,2015

[5] Mohamed A., Luis A.Z., Faisal Z. Q., “Mosaic of near ground UAV videos under parallax effects”, IEEE International Conference on Distributed Smart Cameras, 1-6,2012

[6] Wischounig-Struc D., Quaritsch M., Rinner B., “Prioritized data transmission in airborne camera networks for wide area surveillance and image mosaicking”, IEEE Computer Vision and Pattern Recognition, 17-24,2011

[7] Chris Harris, Mike Stephens, “A combined corner and edge detector”, Proceedings of The Fourth Alvey Vision Conference, 147-151,1988

[8] Jiayuan Li, Qingwu Hu, and Mingyao Ai, “Robust Feature Matching for Remote Sensing Image Registration Based on *Lq*-Estimator”, IEEE Geoscience and Remote Sensing Letters, (99):1-5,2016

[9] Julio Zaragoza, Tat-Jun Chin, Quoc-Huy Tran, Michael S. Brown, David Suter, “As-Projective-As-Possible Image Stitching with Moving DLT”, IEEE Conference on Computer Vision and Pattern Recognition, 36(7):2339-2346,2013

[10] Shu Xin, Li Dong-Xing, Xue Dong-Wei, “Five Frame Difference and Edge Detection of Moving Target Detection”, Computer Systems and Applications, 23(1):124-127,2014

[11] Quanlu Wei, Songyang Lao, Liang Bai, “Long-term Visual Tracking based on Correlation Filters”, AIP Conference Proceedings, Volume 1820,2017