A NEW FRAMEWORK OF MOVING TARGET DETECTION AND TRACKING FOR UAV VIDEO APPLICATION

Wenshuai Yu[a,]*, Xuchu Yu[b], Pengqiang Zhang, Jun Zhou

[a] Institute of Surveying and Mapping,450052, Zhengzhou, Henan, China-ws_yu@yahoo.cn
[b] Institute of Surveying and Mapping,450052, Zhengzhou, Henan, China -xc_yu@yahoo.com

WGS, WG III/5

KEY WORDS: Image Processing, Computer Vision, Motion Compensation, Motion Detection, Object Tracking, Process Modeling, UAV Video

ABSTRACT:

Unmanned aerial vehicle is a new platform for remote sensing, and the primary sensor of it is video camera. Video, also could be called dynamic image is the most important data format which obtained by unmanned aerial vehicle. The combination of video data and UAV provides a novel remote sensing pattern. Moving target detection and tracking is an important technique of video processing for its huge potential in military and other applications. The technique always contains three basic parts: motion compensation, motion detection and object tracking. Each part adopts kinds of technical methods to solve the problems in respective fields. The paper, based on the analysis of the algorithms related to the technology, presents a new framework of it. Different from other moving target detection and tracking frameworks, the framework performs a parallel processing among the three sections by including collaboration control and data capture modules. Comparing with other frameworks, it is more suitable to the UAV applications, because of its advantages such as transferring parameters instead of real data and offering interface to user or exterior system.

1. INTRODUCTION

Unmanned Aerial Vehicle (UAV) is a new developing remote sensing platform, and different from other platforms, for example satellite or airplane, it carries video sensors. So video data is the main information got by UAV. Video could be interpreted as dynamic image, and dissimilar to static image, it can reflect motion information through the changing of gray-level. An important research field of video processing for UAV application is moving target detection and tracking. In actual environment, the moving targets could be vehicles, people or aircrafts, and in some special conditions, these targets might be interesting and valuable. But the problem that detecting the targets from the complicated background and tracking them successively is a tough work.

There many technique methods on moving target detection and tracking. Most of them analysed the problem under the condition of static background, for the stillness of background makes the detection and tracking comparatively easier, and these kinds of method can be used in some applications such as safety monitoring. Contrasting to them, it is much more difficult for target detection and tracking with moving background. Especially for UAV video data whose background changing rapidly and always has complex texture characteristic, it is really a challenging task to solve the technical problem.

For moving target detection and tracking using UAV video, a rather reasonable technical approach is adopted widely. Firstly, in order to compensate the background motion caused by movement of camera, stabilizing the background through the frame-to-frame registration of video image sequence would be taken as a precondition of detection and tracking. A significant product the panoramic image is built in the same process. Secondly, basing on the stabilization of background and employing proper methods, the next operation is separating the target image from the background to realize detection of moving target. Finally, moving target tracking is locating the object in image by means of modeling the target according to target's feature property and choosing appropriate tracking method.

According to the technical approach mentioned above, the technique can be divided into three sections: motion compensation, motion detection and object tracking. It always takes the three parts as a serial course and implements them one after another in a processing. Actually, for there are mutual activities between different sections of the technique, it is not necessary to process the technology orderly, which means executing it step by step. So it not only needs a framework to integrate all these parts, but also requires the framework more effective and practical.

2. MOTION COMPENSATION

Motion compensation is the basic part of the technique, especially for moving background video. It estimates the ego-motion of camera and compensates the background motion of image, and through this way, it makes the moving objects more obvious and the detection of target easier. There are two kinds of approaches adopted, one is feature-based methods, and the other is flow-based methods. Though the latter one has rigorous theory foundation, the former one is more popular. Feature-based methods extract features and match them between image frames to fit the global motion model of video image sequence.

* Corresponding author. Tel.:+86-13526657654; E-mail address:.ws_yu@yahoo.cn.

Feature extraction and matching are prepared for image registration. The image registration that implements frame-to-frame registration of the video image sequence is the key point of motion compensation. The result of image registration could be used in two directions, image stabilization and image mosaicking. Former can restrain the moving background and facilitate the detecting of moving target, and latter can update the local image (always express with the ortho-image) and help to form the trajectory of tracked object.

2.1 Feature Extraction and Matching

In feature extraction, choosing a right kind of feature should be considered for one thing. The feature could be point, line or surface. It has been proven that corner feature is robust and easy to operate. Harris operator (Harris et al., 1988) is a typical corner detector, and its principle is that recognising the features by judging the difference of gray-level's change while moving the search window. Detecting results of two series frames shown in figure 1, and there is good coherence between the two, so it should be thought that the operator has a stable performance and the results could be taken as the input of matching.

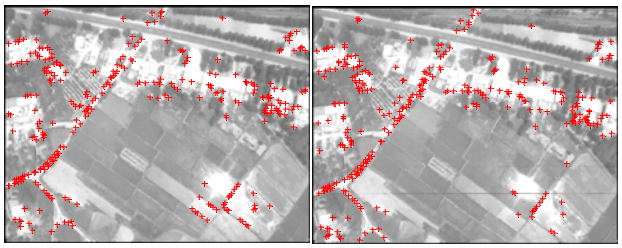

Figure 1. Detecting results using Harris corner operator

After extracting the features, a coarse matching would be made to get approximate matching results, and this course is realized by measuring the similarity of corresponding features. Because there are many mismatches in the approximate results and they cannot meet the requirements of registration, so it has to implement a fine matching to remove the mismatches.

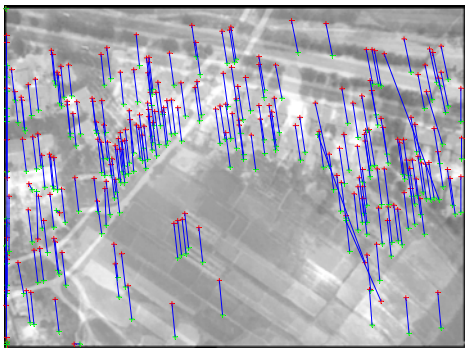

Figure 2. Overlay of two successive frames after eliminating wrong correspondences with RANSAC

A suitable way to keep inliers is combining of epipolar geometry and RANSAC algorithm. Epipolar geometry offers a model—fundamental matrix to the matching, cause the two views should satisfy the epipolar restriction in stereo vision. RANSAC—random sample consensus algorithm (Fischler et al., 1981) is a nonlinear algorithm. Fitting data model with RANSAC maximally restrains the impact of outliers, and reduces the computation to a certain extent. The fine matching is fitting the fundamental matrix through iteration computing and identifying most of the outliers. Figure 2 presents the results of matching after eliminating wrong correspondences from the candidate matches which got from the coarse matching. It can be seen that though bulk of mismatches have been removed, there still a few incorrect correspondences remain.

2.2 Image Stabilization

Image stabilization is compensation of unwanted motion in image sequences. The matter of image stabilization is image registration. The transformation model of image registration is not complicate. A usual choice is affine transformation or projective transformation.

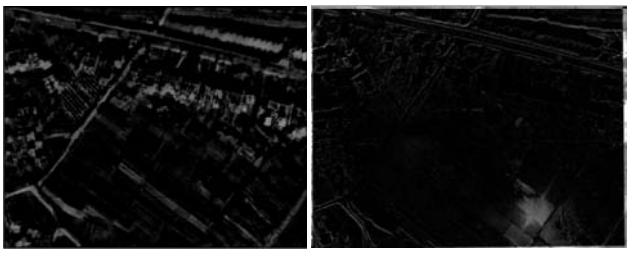

Figure 3. The comparation of difference results before and after image registration

The normal mode for registration is calculating the parameters of the model using corresponding points. Whether the precision of image registration is good or not depends on the results of matching. So image stabilization could be done by computing the registration parameters with the outputs of fine matching and rectifying the prepared frame to reference frame. In order to optimize the result of registration, repeating the course until the accuracy of registration good enough. Figure 3 shows the comparation of difference results before and after image registration. The left one is the difference result previous registration. Except some regions with same textures, most of the background image can not be subtracted, especially some obvious objects and linear features. The right one is the difference result after image registration. Though there are objects edges still distinct, majority of background image got better elimination.

2.3 Image Mosaicking

Mosaicking of video image sequence is rectifying all frames to the reference frame and piecing them together as a panoramic image. The reference frame may be the first frame or a chosen one. A key step for the generation of panorama is image registration.

It is unavoidable accumulate registration errors during aligning the image sequences. The accumulation of errors could induce misalignment of adjoining frames. To resolve the problem, there are many methods have been tried, such as refining registration and introducing reference data. An UAV video image mosaicking is illustrated in figure 4, and there are some piecing seams for registration errors.

Figure 4. A panoramic image mosaiced by UAV video image sequence

3. MOTION DETECTION

The compensation has reduced the impact of background motion, but there are still some influences of it remain in the stabilized image. Motion detection divides the video image into target and background whether it is moving or not. There are many processing methods introduced into motion detection, and the common point of them is the using of motion information. For static background, it usually processes on the background, such as background modeling method. For moving background, it assumes the dynamic image just has target and background two partitions, and if there are more than one target in the video, it will segment the image into numbers of partitions corresponding to the targets, and in some methods it sets the targets on different layers in order to make the process much faster. The primary information for detecting is motion information, or the intensity changes between adjacent video image frames.

3.1 Motion Detection

For video image captured by moving camera, the background motion can't be counteracted absolutely through image stabilization. It may not effective enough to detect the moving target by restraining the movement of background. All the image information could be classified into three kinds: target, background and noise. Different classes correspond to different motion fields in dynamic image. If we know the class characteristics of points, we can use them to fit the parametric sets of different motion regions. Contrarily, if we know the parameters of motion vectors, we could divide the pixels into different fields according motion information. In most of cases, both of the characteristics and parameters are unknown. The clustering of image pixels is a probability question. A typical solution for motion classification is uniting the mixture probability model and EM—Expectation Maximum algorithm (Weiss et al., 1996).

In practice, it can make a hypothesis that there are two layers in the dynamic image, background layer and target layer. After image stabilization, calculating the motion vectors of all pixels and assuming that the flow vectors of target layer is larger than the ones of background layer to estimate the weights of mixture model with iterated computation. It will have the target detected until the iteration convergence. The parameters of image registration could be the initial values of iteration. Figure 5 presents a detection result for one vehicle target in three frames.

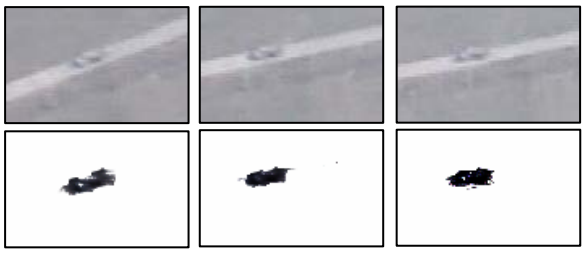

Figure 5. A motion detection result with mixture model and EM

3.2 Motion Segmentation

Motion segmentation is a kind of video segmentation, because it partitions video or image sequence into spatio-temporal regions basing on motion information. Therefore, it is essentially same as the motion detection. Generally, motion segmentation has two basic classes that optical flow segmentation methods and direct methods (Bovik et al., 2005). In perfect cases, there are just two kinds of optical flow associated with the movements of background and target. However, optical flow is not an exact reflection of motion field but an explanation of illumination change. Therefore, it is not rigorous to perform the segmentation with the optical flow information only.

A usually adoption is grouping in motion feature space to realize the segmentation. How to set the relation between clustering and dynamic image is another question. The method of graph theory is a natural solution for motion segmentation. Pixels in image sequence could be taken as the nodes of graph, and if we partition the graph, according motion features, may segment the image at the same time. Edge the weight means the similarity of features between the two nodes which connected by it. In motion segmentation, this similarity measurement is the motion feature vector of each pixel. The graph is not constructed in one image frame. It should connect all the nodes in a spatiotemporal region, and the region may across several frames. After the construction of the weighted graph, it could segment the video image sequence using by normalized cut method (Shi et al., 1998). In order to reduce the complication of computing, an effective solution is subsampling the image sequence by setting spatiotemporal window that just connect the nodes in this window when constructing the weighted graph.

4. OBJECT TRACKING

After detecting the location of target in image, object tracking will persistently lock the position of target during a period. The basic idea of object tracking is modeling the object according to object's feature characteristic property and choosing appropriate tracking method. Different form motion detection emphasizing on accuracy, object tracking couldn't abide taking too much time on computing and needs giving attention on both processing speed and precision, so it has to abstract the target through feature extraction and object modeling. Simply the features used could be shape, size, direction and velocity of the moving object, and complicatedly it could be feature points set, color space and so on. Combining with respective technical approach, it will realize the target tracking. The essence of object modeling is trying to define the target uniquely, and in

single target tracking it only need to depend on one feature property, but in multi-target tracking it may need a integration of different kinds of features for directing at proper target, and it also could using some suitable ways, such as filter methods for multi-target.

4.1 Object Modeling

Object modelling is a representation of object, in other words it utilizes one feature characteristic or the combination of features to express the object. The object's feature could be contour, shape, color, position, texture, velocity and so forth. The more features included, the easier to identify the object. But the combining features will increase burden of processing and demand composite methods. To construct the model of object, we can use the features directly, or transform them into other forms such as templates.

Features of the object may change during the course of tracking, so it requires that the model should be adaptive to the changing or other influences, for example occlusion and unexpected movement. This is considered as the robustness of model. There are many ways to make the model more stable, including using multi-features model and updating the model over time.

4.2 Object Tracking

Using prior information that forms the model of object, tracker predicts the object's position in succedent frames. Corresponding to different models, object tracking has different methods. Object tracking methods attempt to ascertain the coherent relations of feature information between frames, and the strategy of it is no more than searching and matching. Hausdorff distance is a valid measurement for shape and texture features of the object. It can create sparse point sets with feature detectors in images, and the point set of image region labelled as the object is the object's model for Hausdorff measurement. It is able to tackle the deformation of object, because it describes the contour and texture of the object with bulk of points. Taking the measurement and the model, it translates object locating into the matching of point sets (Huttenlocher et al., 1993).

Motion is a kind of state. A typical motion state vector is composed of the object's position, velocity and acceleration along each direction. If the prior and current states are known, the posterior state will be predicted. It is feasible to resolve the problem of object tracking by state estimation means. Kalman filter is one of the state space methods. To define it, the Kalman filter is a batch of mathematic equations that solves the least-squares question recursively. It predicts the values of current state utilizing the estimation values of former state and the observation values of current state, executing the procedure recurrently until the values of every state estimated. To get the estimation values of each state, all the previous observation values have been involved. For object tracking, the state equation is the model of object in Kalman filter, and it describes the transfer of states. The observation is the position of object, and the state vector like mentioned above contains position, velocity and acceleration. Putting the positions of object detected in initial frames into the observation equation of Kalman filter and taking the accurate positions as the initial value of state variant, it compares the output of filtering with precise result to testify the correctness of initial input. It repeats the process until the filter is stable (Forsyth et al., 2003).

Mean-shift algorithm is an approach that searches the maximum of probability density along its gradient direction, as well as an effective method of statistical iteration. Object tracking with Mean-shift algorithm is another class of technique that locates the target by modeling and matching it. Both the modeling and matching are performed in a feature space such as color space and scale space. The mode of it is using the relevant similarity measurement to search the best match. The object tracking basing on Mean-shift algorithm mainly processes on the color feature. Choosing an image region as the reference object model, it will quantize the color feature space, and the bins of the quantized space represent the classes of color feature. Each pixel of the model can corresponds to a class and a bin in the space, and the model can be described by its probability density function in the feature space. Instead of PDF (probability density function), it takes the kernel function as the similarity function to conquer the lost of spatial information. Another reason for using kernel function is smoothing the similarity measurement to ensure the iteration converge to the optimized solution during search (Comaniciu et al., 2003). An object tracking result of airborne video using Mean-shift method is shown in Figure 6.

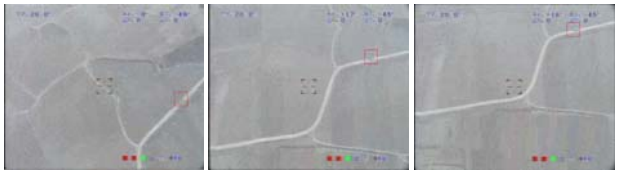

Figure 6. An object tracking result of airborne video using Mean-shift method

5. SYSTEM FRAMEWORK

To the technical approaches analysed above, it needs a framework to integrate all these methods. For the technique of moving target detection and tracking divided into three parts, each part would be an isolated module for its independent function in applicable system. Therefore, the processing is in and between different modules. There are many systems employ a series procedure. Compensation comes first, the next is detection, and tracking put on the last. The reason of that is anterior module always be taken as the precondition of posterior module, and results of each one could be inputs of the next one. However, this kind of system is not considering the interactions between different modules. For example, the result of segmentation can be the initial value of compensation, and the tracking result can accelerate the detection processing.

As shown in the figure 7, distinguishing from traditional technique framework, the presented system framework introduces two more modules, which are data capture and collaboration control. Data capture module gets the video image data and samples it into image sequence, and then it will distribute them to another three modules that are the central parts of the system. The three modules implement a parallel processing, and this will lower the cost of time. After the interior computing, they transfer the outputs that always in the manner of parameters to collaboration control module. The control module manages all the other modules by sending orders to them, and it provides interface to user and exterior system.

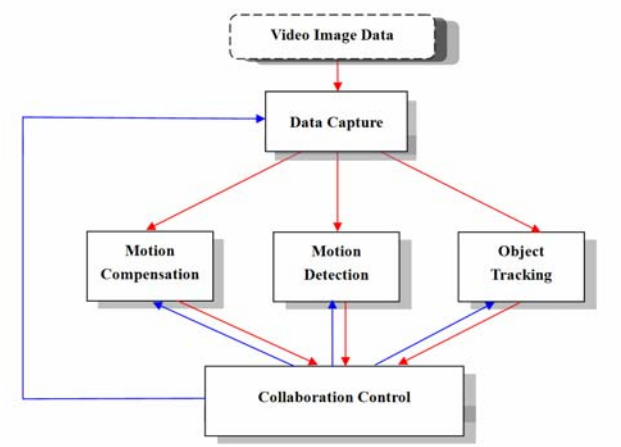

Figure 7. Moving target detection and tracking framework

Figure 8 illustrates the main functional modules of the system. Motion compensation has image mosaicking and image registration two parallel sub-modules. Image mosaicking that could combine with other data mosaics the image sequence, and image registration calculates registration parameters or optical flow vectors. Motion detection includes background subtraction and target detection two serial sub-modules. Background subtraction restrains the movement of background using the parameters or the vectors, and target detection extracts target from the compensated background. Object tracking contains two serial sub-modules that are object modeling and object tracking. Object modelling constructs the model of object with its features. Object tracking realizes the successive locating of the object by utilizing methods corresponding to the model of it.

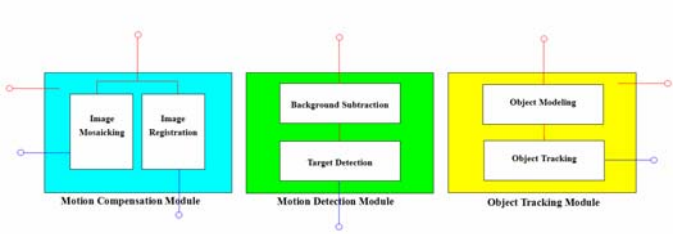

Figure 8. Main functional modules of the system

The advantages of this framework listed as below:

(1) Parallel processing reduces the computation to meet the requirement of real-time application.

(2) Transferring kinds of parameters instead of real data to minimizes the transmission bandwidth.

(3) Users and exterior systems can conduct and monitor the modules through the interfaces offered by control module to evaluate the methods or make improvement.

Moving target detection and tracking is a developing technique, and many technical methods will be invented and introduced for it in future. Though the methods may be diverse in forms and based theories, they have an identical purpose and conform to a regular system framework. Besides integrating the existing methods, another use of this framework is testing the newborn methods.

6. CONCLUSION

On the basis of analyzing the functional parts that motion compensation, motion detection and object tracking and the corresponding technical methods of moving target detection and tracking, we presented a new framework for the technique. We recognize that although there are connections between different sections of the technology, a serial processing of them is dispensable. We realized a parallel computation of the three parts by adding control and capture modules. The design of the framework facilitates the spatial separation of system and reduces the data stream transferred between different modules. This is meaningful to UAV application. Because a typical UAV system composes of aircraft and ground control station, and the data transferring depends on wireless communication.

Our further work includes:

(1) According to the framework, construct the testbed system to test the performance of technical methods and set the standard for evaluation.

(2) Embedding the functional modules into the UAV system and improving them to meet the practical requirements.

REFERENCES

Harris, C., Stephens, M., 1988. *A Combined Corner and Edge Detector. Fourth Alvey Vision Conference*，Manchester UK，pp. 147-151

Fischler, M.A., Bolles, R.C.**,** 1981. *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*. Communications of the ACM, 24(6), pp.381-395

Weiss, Y., Adelson, E.H., 1996. *A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of model*. In Proc. IEEE Conf. on CVPR, pp. 321-326

Alan, B., 2005. *Handbook of Image and Video Processing*. Elsevier, pp. 474-485

Shi, J., Malik, J., 1998. *Motion Segmentation and Tracking Using Normalized Cuts.* Proc. Int'l Conf. Computer Vision, pp. 1154-1160

Huttenlocher, D.P., Noh, J.J., Rucklidge, W.J., 1993. *Comparing Images Using the Hausdorff Distance.* IEEE transactions on Pattern Analysis and Machine Intelligence, 15(9), pp. 850-863

Forsyth, D.A., Ponce, J., 2003. *Computer Vision: A Modern Approach.* Prentice Hall, pp. 380-396

Comaniciu, D., Ramesh, V., Meer, P., 2003. *Kernel-Based Object Tracking.* IEEE transactions on Pattern Analysis and Machine Intelligence, 25(5)：564-577