An Objects Detection Framework in UAV Videos

Jianjiang Lu, Pengfei Fang, and Yulong Tian

Institute of Command Automation,
PLA University of Science and Technology,
Nanjing 210007, China
miipl501@gmail.com

Abstract. Unmanned aerial vehicles equipped with surveillance system have begun to play an increasingly important role in recent years, which has provided a wealth of valuable information for national security and defense system. The automatic understanding technology based on video contents becomes especially important when facing so abundant information. According to the characteristics of UAV videos that moving objects often appear small and background is complex, our thesis makes research among image normalization, histogram equalization, thresholding methods, morphological processing, motion history image and motion segmentation to find out their different effects in foreground detection. What's more, we have designed basic detection method and enhanced detection method in motion objects detection module, which effectively integrates the traditional single-frame detection technology and multi-frame detection technology into our framework.

Keywords: Moving objects detection; UAV; Single-frame detection; Multi-frame detection.

1 Introduction

Objects detection can be divided into static objects detection and moving objects detection, in which static objects detection is achieved by image segmentation and color, grayvalue, edge and texture-based clustering [1] and moving objects detection is achieved by making use of the continuity of moving information to partition areas in images and associate similar moving areas into target objects[2]. Moving objects detection can be divided into three categories including optic flow methods, background differing methods and frame differing methods.

The methods of optic flow need calculating the field of optic flow first which is sensitive to noise and cannot get a precise result, meanwhile, it is not proper to detect real-time moving objects [3]. The methods of background differing find the moving foreground objects by subtracting current frame and background image, and this method can well extract the shape of objects if static background model provided [4]. However, it costs much to rebuild background model for each frame due to the constant moving of cameras. The method of frame differing is the most direct way to detect moving objects, it is widely accepted by many researchers because of its simpleness and efficiency [2, 5], however, there will be some holes in moving objects or the edge of detected entity may be incontinuous around their edges at some time.

M. Zhou and H. Tan (Eds.): CSE 2011, Part II, CCIS 202, pp. 113–119, 2011.

The method of frame-differing can be categorized into single-frame detection technology and multi-frame detection technology according to the number of frames used. Single-frame detection technology detect objects rely on analyzing the difference between objects and background in images mainly achieved by finding a proper threshold to process image after it was filtered. This kind of method often contains morphologic based method, pixel analysis based method, and wavelet based method and so on. When the noise is too much in comparison with the useful information, multi-frame detection technology is needed which mainly utilize the moving objects' continuity and the randomicity of noises to weed out the ostensible targets. In addition, recognition based objects detection technology is coming out in recent years [6, 7].

There are some differences between objects detection in UAV videos based applications and common monitor applications. Most obviously, the moving objects are often small and the targets' intensity is faintness in UAV videos. Nevertheless, researchers have proposed many methods to overcome the difficulty. In paper [8], Saad detect the regions of interest moving objects by extracting object contours and their system is capable of tracking targets as small as 100 pixels. And in paper [9], Andrew obtained their best results by using a combination of intensity thresholding, motion compensation, interest-point detection, correspondence and pattern classification.

Our work is illuminated by above ideas and designed a small objects detection module which contains basic detection module and enhanced detection module in which image normalization, histogram equalization, thresholding methods and morphologic methods contained in based basic detection module and moving history image with motion segmentation contained in the enhanced detection module.

2 Small Moving Objects Detection in Our Framework

The objects detection flow chart we proposed is showed in figure 1 including basic detection module and enhanced detection module. When the shooting environment is not very complex or wicked, morphologic based basic detection module can well exclude the effect caused by noises and find foreground. However, when the illumination changes frequently, cameras often quiver, targets are tiny or objects are moving in a noisy foreground, the contour of moving objects may rupture and blur at some time, in the case, morphologic based methods will not be enough.

In order to overcome the difficulty, we proposed an enhanced detection module that is constituted by motion templates including motion history image and motion segmentation in which motion history image can cumulate the faint changes in each frame and motion segmentation can find the consistent moving foreground [10]. Our basic detection module mainly contains such steps including change detection, image processing and region refining. In our work, change detection gets a differencing image after motion compensation. Image processing gets a foreground mask by extracting objects' contour. Region refining refines objects' area contour through three frame differing.

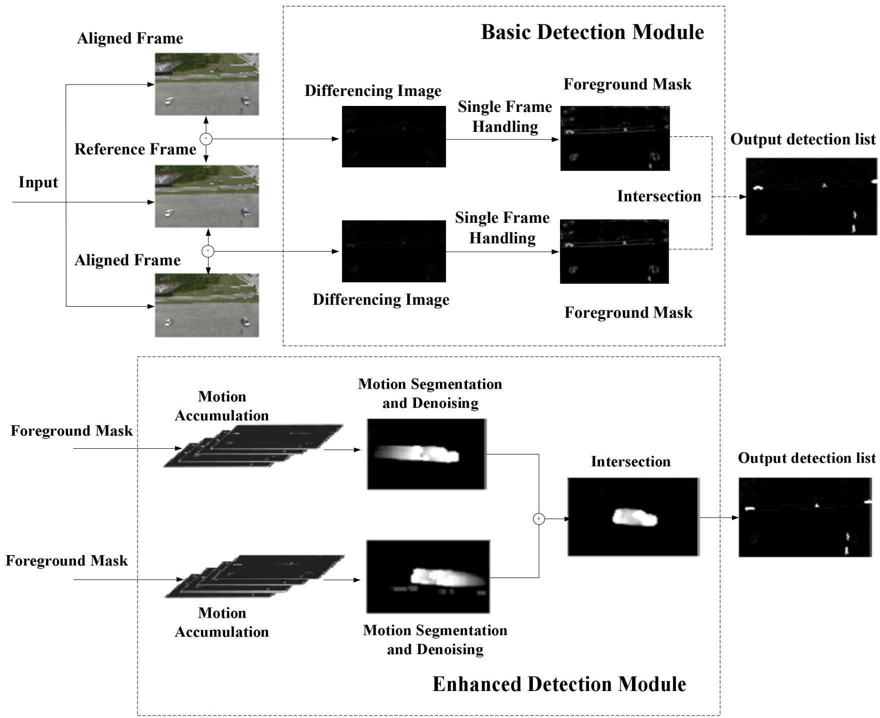

Fig. 1. The framework of our objects detection algorithm

The basic detection module can independently detect the foreground while enhanced detection module can father improve detection capability. When the enhance detection module is needed, the system will not instantly do region refining but process the foreground masks as follow operations including motion accumulation by image history image, motion segmentation to exclude noises and then do region refining at last.

2.1 Change Detection

Change detection finds the difference between two adjacent images by subtracting two aligned images. The input of change detection is current image and reference image, then we align the reference to current image by motion compensation. The output of change detection is a grayvalue matrix constituted by difference gray values from which we can threshold out the moving objects. However, this kind of method needs parallel moving between cameras and targets, and the grayvalue of background keep a constant level in a short time. There are three main approaches in pixel level change detection: background subtraction, inter-frame difference and three-frame difference. However, in the application of UAV video, due to the objects move slowly and the energy of moving foreground is faint, so the method of three-frame difference is not always suitable because it may further lessen the energy of moving foreground.

So we consider if large objects appearing or the velocity of moving foreground is considerable, the effect of three-frame difference is better. Luckily, the forward-backward motion history images help us overcome the problem to some extent because faint energy in each change detection result is accumulated into MHI and it also makes use of the idea of three-frame difference to refine the detection result [5].

2.2 Image Processing

After getting the result matrix by change detection, we need do further processing to find the obvious changing regions and mark them to moving foreground masks. In practice, the masks may contain the moving objects, some static objects and noises which the latter are not needed. A variety of image processing methods are available for us to solve this problem including histogram equalization, image sharpening, normalization, medium/Gaussian filtering, morphological methods and so on. These methods are often applied in different situations.

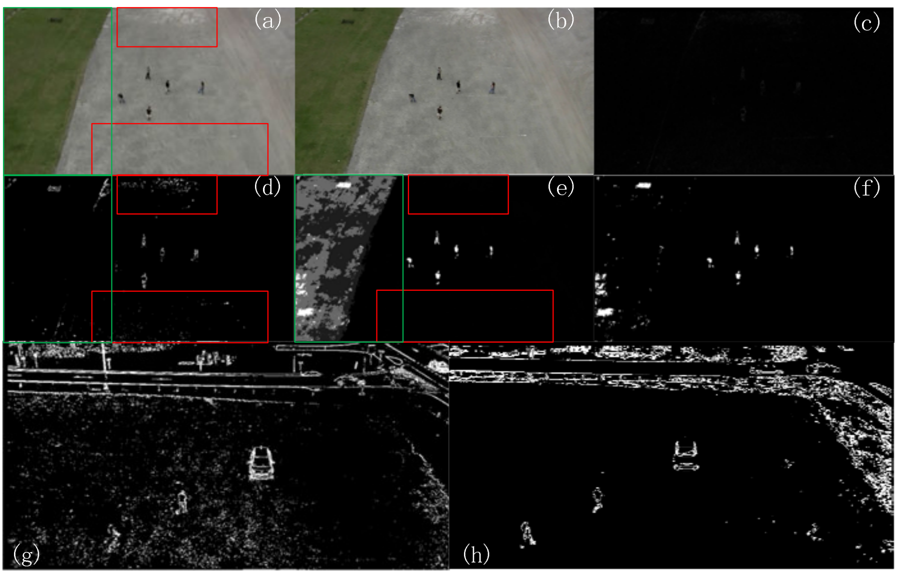

Fig. 2. The effect of normalization to change detection

A notorious problem in airborne video is rapid change in pixel intensities when the camera sensor has automatic gain change control. The gray value of each pixel changes greatly as camera rapidly adjusts its gain to avoid saturation. The changing illumination makes the intensity-based frame difference method inadequate for obtaining accurate motion [5]. So a normalization method is needed and we also take a simple normalization method as forum 1 showed:

$$I'(x, y) = \frac{I(x, y) - \overline{I}(x, y)}{std(I(x, y))} \tag{1}$$

A group of experiments have been done as illustrated in figure 2 in which (a) and (b) represent two adjacent frame, (c) represents the differing result, (d) is the result after differing by the threshold value of 15 in which some pepper-salt noise exist, the method of normalization can weep out this kind of noise but imports some other new noise showed in (e). Figure (g) and (h) is another group of experiment results, (g) stands for not doing normalization for single image before frame-differing and (h) stands for doing normalization for single image before frame-differing. So it is surely the method of normalization can move the pepper-salt noises introduced by illumination. However, when the grayvalue of moving foreground is not obviously different from its around background, the normalization to change detection result may destroy the contour of moving foreground as illustrated in figure 3(d) in comparison with figure 3(c). So it is better to normalize the image by region but not normalize the image as a whole.

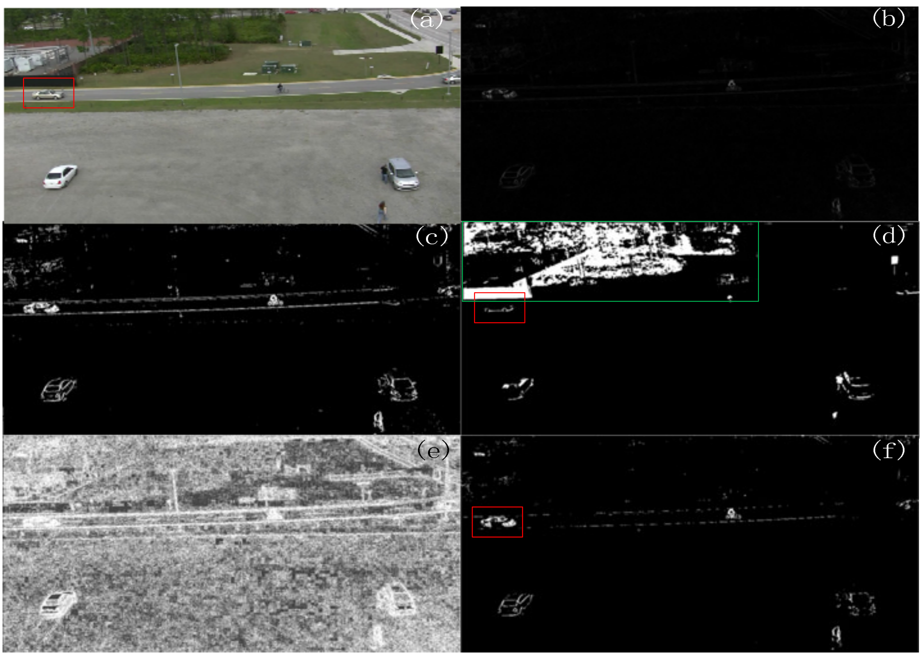

Fig. 3. The effect of white equalization to change detection

Another similar processing method is white equalization also named histogram equalization which is used to make the grayvalue histogram more smoothable. In this paper, we make such steps to achieve white equalization. (1) Count the frequency of each grayvalue appeared in each pixel and calculate:

$$p(i) = \sum_{x=1,\, y=1}^{x=m,\, y=n} \delta[src(x,y) - i] / (m \times n) \tag{2}$$

in which the variable i ranges from 0 to 255. (2)Sum $p(i)$ to form $g(i)$, and $g(i)$ stands for a probability value that grayvalue is less than i and $g(255)$=1. (3)Adjust every pixel by the forum $dst(x,y)=g[src(x,y)]\times 255$ to form the final change detection result. In comparison with the method of normalization, histogram equalization avoid the probability that a large range of noises appearing and can lessen noises at the same time. Other image processing methods about thresholding and morphological methods can reference paper [10, 11].

When the signal to noise rate is very low and random strong noises produce many ostensible targets, single frame detection technology is hard to apply and need multi-frame detection technology to locate moving objects. It mainly includes motion history image and motion segmentation. Motion history image can accumulate faint changes in each differing result [5] and motion segmentation can find consensual moving foreground [10], our enhanced module combine them to overcome the effect of noises and fit for detecting objects of different sizes and velocity.

2.3 Motion History Image

For small objects, because they lack of obvious shape, size and texture information, it's hard to detect moving targets only knowing the foreground mask. On the other hand, the foreground objects are scattered by single frame differing. Motion accumulation [8] or motion history image can be used to get the objects' moving information. In comparison with motion accumulation, MHI not only enhances the foreground, but also keep the information of newest position, and can utilize MHI to do further motion segmentation.

9 19 29

Fig. 4. The detection result using MHI

Motion energy image need to be built before constructing a MHI, Bobick had proposed how to calculate a MEI in paper [12]. However, motion energy image can only present the location of moving objects but cannot describe their moving details. So MHI is also needed. MHI can describe many motion properties such as motion direction. Figure 4 shows the experimental result using MHI to detect moving foreground.

2.4 Motion Segmentation

Motion segmentation is an effective way to track general movement [10]. Although MHI can enhance the faint foreground, it cannot move the effect of noises entirely so we utilize motion segmentation to do further analysis. Motion segmentation aims at calculating every moving foreground blob's global motion vector in its corresponding MHI's lasting time and it relies on motion gradient and stamp information in MHI. Sequentially fading silhouettes record the history of previous movement and thus are referred to as the motion history image. Once the motion template has a collection of object silhouettes overlaid in time, we can derive an indication of overall motion by taking the gradient of the MHI and we can take these gradients by using the Scharr or Sobel gradient functions. Finally, we can collect a measure of global motion.

The longer time the objects last, the more coherent the objects' motion have, the more obvious the interesting moving targets are. At the same time, the noises appear with no regulation and the absolute value of global motion vector in noises is obviously less than that of extracted in moving foreground. So through statistic the global motion vector in each blob, we can further extract interesting targets.

References

1. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530–549 (2004)
2. Richard, J., Srinivas, A., Omar, A.: Image change detection algorithms: a systematic survey. IEEE Transactions on Image Processing 14(3), 294–307 (2005)
3. Sun, Z.H., Zhu, S.A., Zhang, D.W.: Real-Time and automatic segmentation technique for multiple moving objects in video sequence. In: IEEE International Conference on Control and Automation, pp. 825–829 (2007)
4. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. IEEE Transactions Pattern Analysis and Machine Intelligence 22(8), 747–757 (2000)
5. Yin, Z.Z., Collins, R.: Moving object localization in thermal imagery by forward-backward MHI. In: Proceedings of IEEE Computer Science Conference on Computer Vision and Pattern Recognition, pp. 133–140 (2006)
6. Bai, X., Wang, X., Latecki, L.: Active Skeleton for Non-rigid Object Detection. In: International Conference on Computer Vision (2009)
7. Maji, S., Malik, J.: Object Detection using a Max-Margin Hough Transform. In: Proceedings of IEEE Comference on Computer Vision and Pattern Recognition, pp. 1038–1045 (2009)
8. Ali, S., Shab, M.: COCOA - Tracking in aerial imagery. In: Proceedings of SPIE Airborne Intelligence, Surveillance, Reconnaissance(ISR) Systems and Applications, Orlando (2006)
9. Miller, A., Babenko, P., Hu, M.: Person tracking in UAV video. In: Proceedings of International Workshop on Multimodal Technologies for Perception of Humans, Berlin, Heidelberg, pp. 215–220 (2008)
10. Bradski, G., Kaebler, A.: Learning OpenCV. O' Reilly Media, Inc., Sebastopol (2009)
11. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Transactions on Systems Man and Cybernetics 9, 747–757 (1979)
12. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 3(3), 257–267 (2001)