ISBN: 978-972-8924-63-8 © 2008 IADIS

MOTION SEGMENTATION AND MEANSHIFT ASSISTED CONTOUR REFINEMENT FOR AIRBORNE VIDEO

Ratheesh K
Cochin University of Science and Technology

Supriya Rao
Honeywell Technology Solutions Lab

Santhosh Kumar G
Cochin University of Science and Technology

ABSTRACT

This paper presents methods for moving object detection in airborne video surveillance. The motion segmentation in the above scenario is usually difficult because of small size of the object, motion of camera, and inconsistency in detected object shape etc. Here we present a motion segmentation system for moving camera video, based on background subtraction. An adaptive background building is used to take advantage of creation of background based on most recent frame. Our proposed system suggests CPU efficient alternative for conventional batch processing based background subtraction systems. We further refine the segmented motion by meanshift based mode association.

KEYWORDS

Machine vision, motion segmentation, meanshift, airborne video, surveillance, and unmanned aerial vehicle.

1. INTRODUCTION

There has been a great deal of interest in computer aided processing of vast amounts of aerial video data in real time or in near real time scenarios. Moving object detection and tracking in static camera is a well researched area, with a lot of reliable techniques in literature. In the case of an airborne observer the background as well as the objects of interest moves independently. If the motion of observer is significant, causing a less overlap of consecutive frames, most techniques adopted from the static camera scenario will become unsuccessful. Also, the quality of object detection has significant impact on the later phases of surveillance such as tracking, trajectory documentation, etc. Airborne videos are usually characterized by small target size, cluttered background, low contrast, camera jitter other than ego motion, etc. In such case, tracking of the objects needs a discriminative object representation and an efficient computation for a real-time performance.

Segmentation of moving objects from a moving camera is addressed by Burt (1989), Rosenberg (1998), etc. In particular, motion segmentation for an airborne video is discussed by Wildes (2001) and Cohen (1998), etc in greater details. Air borne video is less affected by depth variations and parallax in the background when compared to the video taken from a ground vehicle. The separation of object motion from ego motion can be achieved by an optical flow (Cohen, 1996), temporal gradient (Halevi, 1997) and correspondence based methods (Irani, 1998), etc. Background based methods outperforms all these methods when camera motion is nil or ignorable. Such constraint allows a background to be evolved from sufficient number of frames with considerable overlap and gives quality segmentation of moving objects. A median background based method proposed by Ronald (2006), require a window of frames to be stored in memory with batch processing model of entire window.

Motion segmentation often suffers from the poor foreground/background classification near boundary region. High quality contour identification is an important prerequisite for feature based object tracking systems. Density gradient estimation procedure, the mean shift in a spatial-value domain of an image is a

good choice for autonomous static image segmentation (Comaniciu, 1999). Combining the high level static image segmentation information and local motion cues can improve the quality of object detection.

Our major contributions through this paper include a background model for subtraction based object detection, and a meanshift based refinement of contour for quality object segmentation.

2. MOVING OBJECT DETECTION

Usually, the moving objects like cars and other vehicles need to be detected for surveillance purpose. Keeping motion as primary cue, the objects are detected by change analysis. While motion detection techniques are fairly robust in the case of static cameras, there are challenges in adapting them to the moving cameras. Often, the first step is to compensate for the camera motion (Ego motion) by image registration. Later the set of registered frames can be used to model background.

2.1 Ego Motion Estimation

The detection of moving objects in case of a moving camera is complicated by the movement of camera as well as object. To filter out the object motion, we need robust camera motion estimation. Modeling the motion as 2D affine or projective transforms, assuming the background is roughly planar, gives reasonable accuracies. Under an affine model, pixel locations $\mathrm{x} = (x,y)$ in frames I^{n} and $\mathrm{I}^{\mathrm{n}+m}$ are related by the transformation

$$\mathrm{x}_{\mathrm{n+m}} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (1)$$

Estimation of affine matrix needs minimum three set of corresponding points in I^{n} and $\mathrm{I}^{\mathrm{n}+m}$. There are several techniques like a good feature selection (Shi and Tomasi, 1994), scale invariant feature transforms (Lowe, 2004), etc that can be used to locate easily traceable feature points. These can then be located in subsequent views for registration. The registered point set will usually include outliers, as the image contains moving regions too. Outliers are typically removed from the correspondence set by using **RANSAC** algorithm (Fischler, 1981). This gives robust registration of images and hence the measurement of camera motion.

2.2 Background Building Problem

In motion based object detection, the best results are derived using background subtraction approach. In this section we discuss the existing background modeling methods and their limitations.

For a static camera video, Gaussian background model is reasonable for characterizing the pixel intensity variation. In such cases, the effective overlap of frames allows a good model to evolve over time. Here, in airborne video, the camera motion prevents the stacking of registered frames for enough duration. In Mixture of Gaussian model, each pixel is modeled separately (Chris, 1999) by a mixture of K Gaussians:

$$\mathrm{p}(\mathrm{I}_\mathrm{t}) = \sum_{\mathrm{i}=1}^{\mathrm{K}} \omega_{\mathrm{i,t}} \eta(\mathrm{I}_\mathrm{t}; \mu_{\mathrm{i,t}}, \sum_{\mathrm{i,t}}) \quad (2)$$

Where K is usually predefined as 3 or 5. According to the range of I_t background, components are updated. The component with high evidence and low variance is selected for subtraction and thresholded to extract foreground. Here the memory requirements are limited but it takes more computation time. This background model is only applicable when the camera movement is negligible.

On the other hand the “double difference” method suggested by Kameda (1996) gives poor results on slow moving and low textured objects due to foreground aperture problem as shown in figure 1. In this method thresholded difference between frames at time t and t − 1 and between frames at time t−1 and t−2 are calculated. Combining the difference images with a logical AND gives the motion segmentation. Here the differencing is done after all frames are registered to frame at time t.

ISBN: 978-972-8924-63-8 © 2008 IADIS

Ronald (2006), suggests background building method for air borne moving camera using median of registered frame stack. This method benefits from fast bootstrapping and good quality of object detection. But the algorithm demands inherent memory storage of full frame stack and variable quality of detection imposed by far and near frame registration over the window. Moreover the window based batch processing is not acceptable for real time tracking and precision targeting systems, which is a strong use case in airborne surveillance applications.

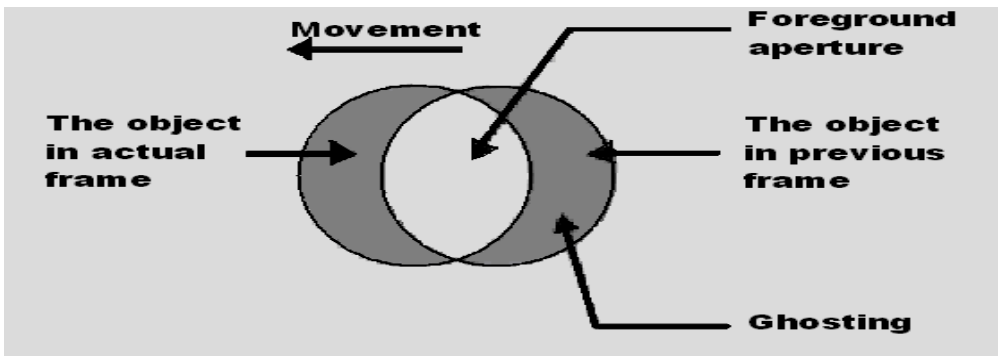

Figure 1. Performance drawbacks of double difference algorithm on a uni-colourd object. Holes are created if the movement of object is not fast enough.

2.3 Proposed Background Model

Here we propose a background model, which gives reasonable quality of detection with less memory requirements and processing power.

In order to carry out background subtraction, both frames and background needs to be registered. The motion compensation transformation functions such as affine or projective requires interpolation steps which corrupt the image. The median filtering on such a set of registered frames imposes a smoothing effect as shown in Figure 2B. Owing to this smoothing effect, often, it is difficult to locate sufficient number of feature points for a real time updation of background by registering to the current frame.

Here we propose an algorithm for moving object detection where the smoothing effect is avoided. Additionally, it also simplifies the memory requirement by not requiring a full memory stack of all previous frames.

In our algorithm the bootstrapping is done by a median filter over N initial frames, as described by Ronald (2006).

Let $I_1 I_2 I_3 \ldots I_N$ be a stack of N initial frames which are registered to $I_{N/2}$. The initial background frame B_{In} is calculated as:

$$B_{In}(x, y) = \text{median}\{I_i(x, y), i = 1..N\} \quad (3)$$

This background can be used for all registered frames in the window $I_1 \; to \ldots I_N$. But the quality of object detection will vary as the registration anomalies will be more for the frames with less overlap. Another possibility is to repeat the frame stacking for each frame, creating its own background. But this will demand high processing power of N registrations for each frame.

In our case, after the initial boot strapping, we update the background information instead of re computing it. Owing to this, there is no need to keep the stack of registered frames in memory. Let Thr be the predefined threshold and α the predefined value between 1to 0 for alpha blending. The updation of the background is as follows:

$$IF\left(B_t(x, y) - I_t(x, y) \geq Thr\right)$$
$$B_{t+1}(x, y) = \alpha * I_t(x, y) + (1 - \alpha) * B_t(x, y)$$

ELSE

$$B_{t+1}(x, y) = I_t(x, y) \quad (4)$$

$$\text{Register}(B_{t+1}, I_{t+1}) \quad (5)$$

As described in equation 4, background B_{t+1} for an image I_{t+1} at time t+1 is calculated differently for pixels near object motion area and no motion area. An Alfa update is used to adapt the background changes near motion area. On the other hand, most of the unaffected areas are preserved without doing any pixel manipulation operations. As a result the smoothing effect is localized. A Register () function, as described in the equation 5 registers the background image to current image as discussed section 2.1. Here we make use of most part of recent frame for background building. The background builds sharp enough in the area where the object movement is not found. This helps for further registration process. We can build quality background with less memory requirement and processing power compared to other methods.

An example of the generated background and corresponding object detection is shown in figure 2C and 2F respectively. The quality of object detection is on par with the median background based systems, despite the low memory and processing requirements. As a further refinement, in section 3 we will demonstrate how to handle the subtraction anomalies by mean shift segmentation based method.

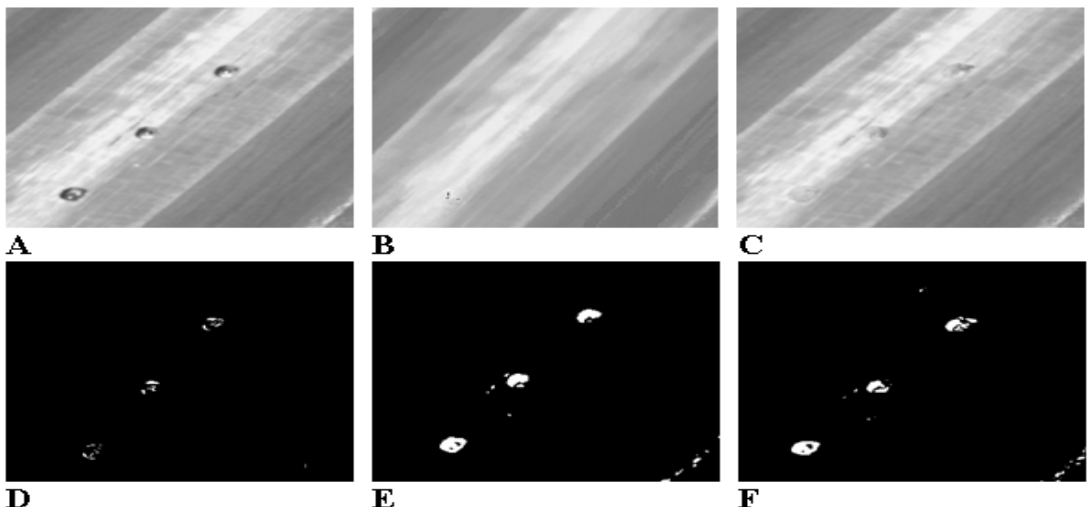

Figure 2. Object detection comparison .A shows the original image, B median background, C background using proposed method , D the blobs using double differencing, E the blobs using median method, and F blobs using proposed method.

3. MEANSHIFT BASED CONTOUR REFINEMENT

The Meanshift technique is a method to locate the stationary points of a distribution and in turn locate the modes of the distribution. Image Segmentation algorithms based on Meanshift procedure have been discussed in the literature with very promising results (Christoudias, 2002) and (Comaniciu, 1999). The advantage of Meanshift segmentation is that they are controlled by a very few tuning parameters and hence more autonomous in nature. Meanshift clustering does not require the knowledge of the number of clusters and have no constrain on shape of the clusters. A brief introduction to the mean shift procedure and explanation of how to use the mean shift based clustering to improve the object detection performance are given below.

3.1 Meanshift Procedure Overview

ISBN: 978-972-8924-63-8 © 2008 IADIS

Given n data points $x_1 \ldots x_n$ in the d-dimensional space Rd the kernel density estimator with kernel function K(x) and a window bandwidth *h*, is given by Comaniciu (2002).

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right) \tag{6}$$

Where the *d*-variate kernel *K*(x) is nonnegative and integrates to one. Using a radially symmetric kernel $K(\mathrm{x}) = \mathrm{c}_{\mathrm{k,d}}\mathrm{k}(\| \mathrm{x} \|^2)$ the density estimator can be written as:

$$\hat{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\| \frac{x - x_i}{h} \|^2 \right) \tag{7}$$

Standard mean shift algorithm is a steepest ascent procedure which requires estimation of the density gradient. Selecting Epanechnikov as the kernel of choice it can be showed that the meanshift vector is:

$$\mathrm{m}(\mathrm{x}) = \frac{\sum_{i=1}^{n} x_i \, \mathrm{g}\left(\| \frac{x - x_i}{h} \|^2 \right)}{\sum_{i=1}^{n} \mathrm{g}\left(\| \frac{x - x_i}{h} \|^2 \right)} - \mathrm{x} \tag{8}$$

We start with an initial point and by using the mean shift updates given by equation 8, the algorithm eventually converges to a stationary point. Using different initial points, it is possible to locate all the modes of the underlying density. Once the modes are located, the association of the individual data points to the modes can be done.

3.2 Contour Refinement

The background subtraction based object detection relay on the pixel to pixel colour difference for foreground identification. Thresholding out the foreground becomes tricky, in case it has a similar colour to the background. Moreover, poor contrast near edges of moving objects is a usual scenario in airborne videos. It will be beneficial to aid the decision process by confidence maps derived based on the high level information of region continuities and breaks.

The Meanshift clustering pixels in a colour and range feature space provides high quality image segmentation results. In our implementation, Meanshift procedure was run over the 2D image in LUV color space to locate the clusters in the image. After a mode fusion, the elimination of very large clusters is done as a post processing step. Large regions in an airborne video usually refer the uniform background in the scene. Figure 3 shows an example of the input image and its corresponding mean shift segmentation. Each pixel now can be assigned a label indicating which mode it corresponds to.

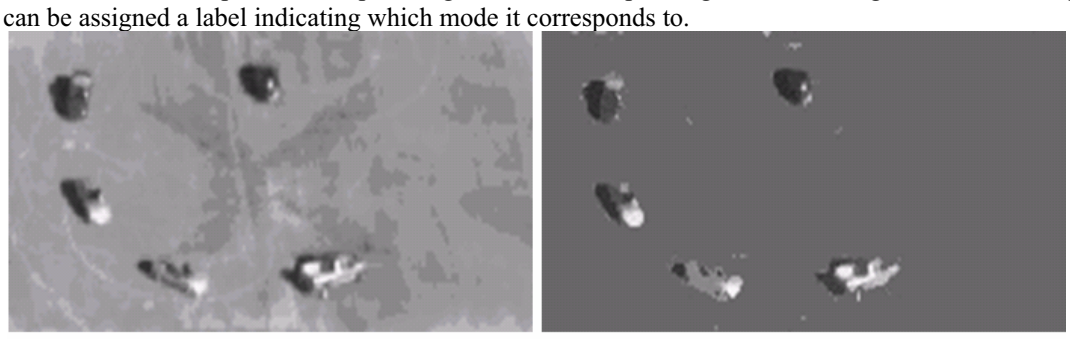

Figure 3 Meanshift image segmentation. A shows the original image sample taken from a UAV, B shows the corresponding segmented image with mode fusion and post processing.

Additionally, as part of the earlier motion detection procedure, each pixel already has a label indicating whether it is part of foreground or background. Depending on the relative position of foreground and

background pixels, each point is assigned a weight to match its confidence of classification. A foreground pixel surrounded by more similar pixels is weighted more as foreground pixel. Similar is the case for background pixels. Based on the assumption that a cluster of pixels as a result of meanshift segmentation belongs to the foreground or background as whole, the foreground refinement process is done as shown in algorithm 1.

```
  Input -Initial contour
        -Mode table from meanshift segmentation
BEGIN
    Select each pixel from image
        If pixel belong to initial foreground
                Add weighted vote to corresponding mode
        Else
            Minus weighted vote from corresponding
                mode
    End of select
    Select each Mode from Mode Table
        If mode vote is positive
                Add entire cluster to foreground
        Else
                 Delete entire cluster from foreground
    End of select
END
```

Algorithm 1. The motion segmentation refinement using meanshift clustering.

As a result of the above stated procedure, the motion segmented objects gets much clear blob with more meaningful contours. Figure 4D shows a typical application of contour refinement. In the figure, a car is motion segmented to multiple parts including miss classified regions. A meanshift clustering augmented with the weighted votes from initial contour refines the final contour.

Note that this algorithm does not demand that the mean shift technique perfectly segment the background and foreground. It works just as well when the foreground and background are over-segmented as several clusters. This is useful in setting a threshold for the meanshift clustering. Typically, the uniform background areas are segmented as a big block while the foreground is segmented into several smaller regions.

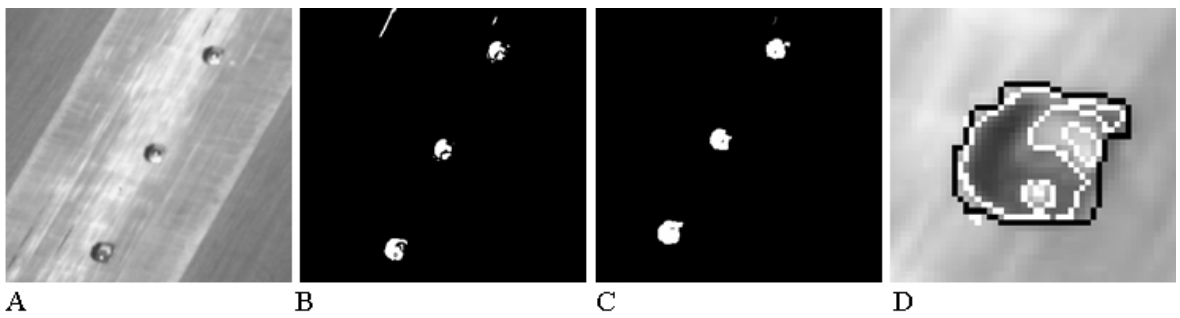

Figure 4. The figure 4 A shows the original image, B initial blobs input, C blobs with refined contour, D the plot of initial(White) contour and refined contour(Black) on zoomed image portion.

4. CONCLUSION

ISBN: 978-972-8924-63-8 © 2008 IADIS

In this paper we discussed alternate methods for moving object detection from airborne video. The approaches presented here serves as computationally feasible methods for low end systems. In our experiments we observed that quality of object detection has dramatic improvement when the suggested background subtraction method is combined with meanshift based edge refinement. A correlogram based tracker in the tracking module works well enough with the stated object detection module.

The above described object detection suffers from motion status change of objects. Autonomous alpha adapting methods need to be developed for minimizing user interactions.

Segmentation logic has its on disadvantages as a top down method. Incorporation of continues motion information for further meaningful segmentation can be tried as a future work. Different data association methods can further improve the segmentation and object detection.

REFERENCES

Benjamin Deutsch. et al, 2005, A Comparative Evaluation of Template and Histogram Based 2D Tracking Algorithms. *DAGM-Symposium* pp. 269-276.

Burt, P. et al, 1989, Object tracking with a moving camera. *Workshop on Visual Motion.* Irvine, CA, USA, pp. 2-12.

Chris Stauffer. and Grimson, W.E.L., 1999 Adaptive background mixture models for real-time tracking. *Proceedings of CVPR '99*, pp. 246–252.

Christoudias, C. et al, 2002, Synergism in low level vision. In *International Conference on Pattern Recognition*, Quebec City, Canada.

Cohen I. and Herlin, I., 1996, Optical flow and phase portrait methods for environmental satellite image sequences. *ECCV*, Cambridge, USA, pp.141-150.

Cohen, I. and Medioni, G., 1998, Detecting and tracking moving objects in video from an airborne observer. *DARPA Image Understanding Workshop,* IUW98, Monterey.

Comaniciu, D. and Meer, P., 1999, Mean Shift Analysis and Applications. *ICCV*. , Corfu, Greece, pp. 1197–1203.

Comaniciu, D. and Meer, P., 2002, Mean shift: A robust approach toward feature space analysis.*IEEE Trans. Pattern Anal.Machine Intell.*, 24. pp. 603–619.

Davide A. et al, 2006, A revaluation of frame difference in fast and robust motion detection, *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, VSSN06 pp. 215-218 .

Fischler, M.A. and Bolles, R.C., 1981, Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM*, 24(6) pp. 381–395.

Halevi, G. and Weinshall, D., 1997, Motion disturbance: Detection and tracking of multi-body non rigid motion. *CVPR, IEEE*, Puerto-Rico, pp. 897-92.

Irani, M. and Anandan, P., 1998, A Unified Approach to Moving Object Detection in 2D and 3D Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6).

Kameda, Y. and Minoh, M., 1996, A human motion estimation method using 3-successive viedo frames. *ICVSM*, pp. 135–140.

Lowe, D., 2004, Distinctive image features from scale-invariant keypoints. *Intl. J. ofComp. Vision* 60 pp. 91–110.

Lucas, B. and Kanade, T., 1981, An iterative image registration technique with an application to stereo vision. *Proceedings of DARPA Image Understanding Workshop*, pages pp. 121–130.

Ronald, Jones. et al, 2006, Video Moving Target Indication in the Analysts' Detection Support System. , *Intelligence, Surveillance and Reconnaissance Division Edinburgh*, S. Aust.: DSTO.

Rosenberg, Y. and Werman, M., 1998, Real-Time Object Tracking from a Moving Video Camera: A Software Approach on a PC. Proc. of *IEEE Workshop on Applications of Computer Vision*, pp. 238–239.

Shi, J. and Tomasi, C., 1994, Good features to track. Proc. *IEEE International Conf. Computer Vision and Pattern Recognition (CVPR)*. IEEE Press.

Wildes, R. et al, 2001, Aerial video surveillance and exploitation. *Proceedings of the IEEE, vol. 89*, no. 10, pp.1518--1539.