Video Objects Detection Using Spatial and Temporal Segmentation

Hima E P

Department of Computer Science, University of Calicut, Kerala, India

Abstract: ***Moving object detection and tracking in a Video sequence is a crucial task in many computer vision and image analysis applications such as video surveillance, event detection, activity recognition etc. I here proposed a new method for detecting video objects. This method includes two schemes. These are Spatio-temporal spatial segmentation and temporal segmentation. Combinations of these two schemes are used to detect the moving objects. Both single and multiple objects could be detected through this scheme. Here Markov Random Field (MRF) model is used to represent image. It can be used to model spatial constraints such as smoothness of image regions, spatial regularity of textures and depth continuity in stereo construction.***

Keywords: Image segmentation, edge detection, Markov random field, Change detection mask, Maximum a posteriori probability**.**

1. Introduction

Videos can be defined as sequences of images, each of which called a frame, displayed in fast enough frequency so that human eyes can percept the continuity of its content. It is obvious that all image processing techniques can be applied to individual frames. Besides, the contents of two consecutive frames are usually closely related .So, by analyzing these similarities of the consecutive frames we can find out the moving objects in a video sequence. Moving object detection in a video is the process of identifying different object regions which are moving with respect to the background. In this paper, I propose a compound MRF model based scheme that detects moving objects with less computational burden. This method can be used to track moving objects in the absence of any reference frame, and it can be also useful to detect the objects when they are moving very slowly or do not have much movements. This scheme includes both spatiotemporal spatial segmentation and temporal segmentation. Here, we obtain spatio-temporal spatial segmentation first for a given initial frame by edge based compound MRF Model. Thereafter, for subsequent frames, segmentation is obtained by adding some change information of these frames with initial frame segmentation result. Here spatial segmentation helps in determining the boundary of the regions in the scene more correctly, and temporal segmentation helps in determining the foreground and the background parts of it.

2. Related Works

Although has been studied for dozens of years, object detection and tracking remains an open research problem. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how you define the object to be detected and tracked. If only a few visual features, such as a specific colour, are used as representation of an object, it is fairly easy to identify all pixels with same colour as the object. On the other extremity, the face of a specific person, which full of perceptual details and interfering information such as different poses and illumination, is very hard to be accurately detected, recognized and tracked. As a first stage of study process, I go through various image segmentation methods for different types of images [1] .And Different stochastic model based approaches [2] are also available in the literature and they provide better results. MRF model, because of its attribute to model spatial dependency, is proved to be a better model for image segmentation [4].Both MRF and Hidden MRF models [5], [6] have also been used for moving object detection for the last two decades. Since in a video, spatial and temporal coherence is there, MRF model is shown to be a better resilience .Here image frames are represented using MRFs and spatial segmentation problem has been formulated in spatiotemporal framework or environment. The spatio-temporal spatial segmentation thus obtained is combined with the results of temporal segmentation to detect the moving objects. For temporal segmentation, several change detection mask methods [8] can be used. Image thresholding methods [3] also have their importance.

3. Proposed Scheme

This proposed scheme includes two methods. First one is spatio-temporal spatial segmentation and the other one is temporal segmentation. Spatial segmentation means partitioning the image into a number of arbitrarily shaped regions, each of them typically being assumed to constitute a meaningful part of the image. Temporal video segmentation aims to partition the video elementary image sequence termed scenes and shots. A shot is defined as a set of consecutive frames taken without interruption by a single camera. A scene may contain one or more shots.

Here spatial segmentation is considered in spatio-temporal environment. Hence we use the name spatio-temporal spatial segmentation. And also here we consider the gray value of the images. So the gray value in spatial direction, gray value in temporal direction and edge line field in both spatial and temporal directions are modeled with MRF model. An edge based segmentation scheme was used in initial frame segmentation. For this purpose we can employ either sobel operator or canny operators according to the nature of the image sequences. In temporal segmentation change detection mask (CDM) is used. It is obtained by taking the difference between the two consecutive frames, where information from the previous frame is fed back and the label of the current spatial segmentation result is used to modify the CDM. The modified CDM itself represents a binary mask of the foreground and background region where VOP is extracted by superimposing the original pixels of the current frame on the foreground part of the temporal segmentation.

4. Spatial Segmentation

A compound MRF model is used to model each image frame. And the segmentation problem is solved using the MAP estimation principle. A hybrid algorithm is used to get initial frame segmentation. For the segmentation of other frames, changes between the frames are imposed on the previously available segmented frame so as to have an initialization to find the segmentation result of other frames.

Here the segmentation problem can be considered as Markovian labeling process. And it is guided by maximum a posteriori (MAP) criterion .In object segmentation, only the foreground label F and background label B are used.

Let S be the image and L be the label set. L consists of B and F. The intensity of pixel i is denoted by xi and its label by fi. To find the best segmentation label for pixel i, the probability P (fi |xi, F-{fi}) should be maximized. Where F-{fi} is the segmentation of image except for the i-th pixel .For doing this, we can use the baye's theorem:

$$P(fi \mid xi, F-\{fi\}) = \frac{P(xi \mid fi, F-\{fi\}).P(fi \mid F-\{fi\})}{P(xi \mid F-\{fi\})}$$

The intensity of the i-th pixel (xi) only depends on fi and is independent of other regions F-{fi}.In addition P (xi) has a defined value and does not depend on segmentation. So to maximize the previous equation, only the numerator needs to be maximized.

$$P(fi \mid xi, F-\{fi\}) = P(xi \mid fi).P(fi \mid F-\{fi\})$$

For general segmentation, it is assumed that each label type has a Gaussian distribution with the mean μ and variance σ2.So we have

$$P(xi \mid fi) = 1/\sqrt{2\Pi}\sigma \exp[-1/2\sigma^2 (xi- \mu fi)^2]$$

In which μfi is the mean intensity of region fi. For the whole image, the segmentation problem is reduced to the minimization of following energy function

$$U(S, F) = \sum_{I \in S} U(xi, fi)$$

4.1 Spatial Segmentation of Initial Frame

For the segmentation of the initial frame a hybrid algorithm can be used. For example, we can employ a combination of simulated annealing (SA) and iterated conditional mode. (ICM)

4.2 Spatial Segmentation Of Subsequent Frames

Here we can use change information. It is obtained by computing absolute values of the intensity difference between the current and the previously considered frames followed by a thresholding approach. The pixel values of the changed region of the current frame are superimposed on the previously available segmented frame to get an initialization.

Let *pt.* denote a frame at time *t*, whose spatio-temporal Spatial segmentation q_t is available with us. Now considering p_{t+d} as a frame at an instant (t + d), q $_{(t+d)i}$ represents its initialization obtained by this scheme. qx_{t+d} represents its final spatio-temporal spatial segmentation. x (t+d) i can be obtained as follows.

1)Changed region corresponding to the frame pt+d is obtained by taking a difference of the gray values of the frames pt+d and pt followed by thresholding, and the changes thus obtained is denoted by p(t+d)|pt+d−pt |.
2) The pixels corresponding to these changed regions in the t-th frame segmentation result *qt* are initialized as qtt = qt − qt|pt+d−pt |.

These regions in the qtt are replaced by the original gray values of *pt+d* for initialization of the (t + d) th frame as q (t+d) i =qtt + p (t+d)|qt+d−pt
| p (t+d) |pt+dpt | represents the pixels of the (t + d)th frame where changes took place from the previous frame. q (t+d) i serve as the initialization for spatio-temporal spatial segmentation of the (t + d) th frame. ICM is run on the (t+d) th frame starting from q (t+d) i to obtain q(t+d).

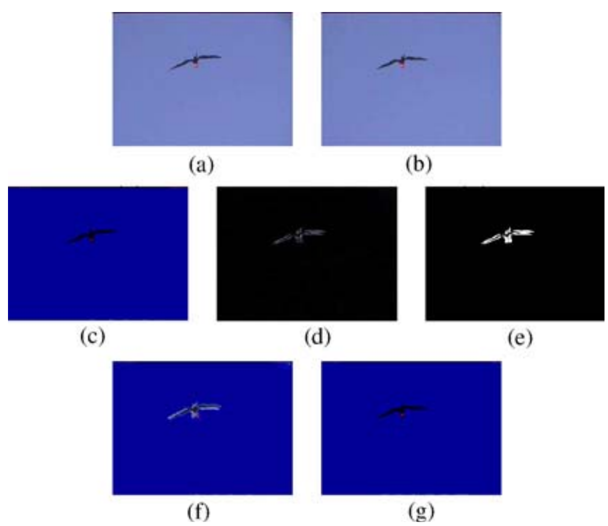

Figure 1: Bird video. (a) Original frame 30. (b) Original frame 34. (c) Edge based segmentation result of 30th frame. (d) Difference image obtained by pixel by pixel comparison of 30th and 34th frames. (e) Threshold difference image for 34th frame. (f) Initialization for 34th frame. (g) Final segmentation result of 34th frame

5. Temporal Segmentation

By doing temporal segmentation, we can classify the foreground and the background in a video image frame. A conventional approach for the detection of changes in images is to pair wise subtracting successive images in the sequence. Change detection methods segments each frame into two regions namely changed and unchanged regions in case of a static camera or global and local motion regions in case of a moving camera. Spatio-temporal change detection deals with the former case where unchanged region corresponds to background and changed regions to the foreground object(s). There are various methods of video object segmentation, but the faster video object segmentation techniques are based on change detection approach followed by further post processing. Change detection can be performed in two ways. First one is pixel based and the second by block based. I here used block based change detection. It can be performed in order to increase the quality of change detection but also to decrease the computation time. Use of blocks allows a processing which is intermediate, between local level like pixel based methods and global level as histogram based methods. Main advantage of this method is their relative insensitivity to noise and camera or object motion. I have adopted a popular thresholding method called Otsu's method for thresholding the CDM. The results thus obtained are verified and compensated by considering the information of the pixels belonging to the objects in the previous frame, to improve the segmentation result of the moving objects. This is represented as

$$S = \{s_{i,j} \mid 0 \le i \le (M-1),\ 0 \le j \le (N-1)\}$$

Where S is a matrix having the same size of the frame. $s_{i,j}$ is the value of the VOP at location (i, j), where (i, j) location represents the i- th row and the j-th column. If a pixel is having a value $s_{i,j} = 1$, then it belongs to a moving object in the previous frame; otherwise it belongs to the background in the previous frame. According to this information, CDM is modified as follows: if it belongs to a moving object in the previous frame and its label obtained by spatio-temporal spatial segmentation is the same as that of one of the corresponding pixels in the previous frame, then it is marked as a foreground area in the current frame else as the background. The modified CDM thus represents the final temporal segmentation. Here CDM is not only generated from the frame difference of current frame and previous frame but also from the frame difference between current frame and background frame, which can be produced by background registration technique. Since the background is stationary, it is well-behaved and more reliable than previous frame.

6. Video Object Plane Generation

After doing temporal segmentation, now we have two classes. One class, which contains foreground parts and the second class, contains only the background, still parts. The class having foreground part is identified as moving object part. And the pixels belonging to that region in the original frame will form the VOP.

7. Experiment and Results

Here the spatial segmentation of the initial frame is done using an *edge based* compound MRF model followed by the hybrid algorithm for MAP estimation. Change information based initialization scheme is used for the spatial segmentation of the subsequent frames. CDM based method is used to obtain temporal segmentation. And the combination of these two was used to generate VOP. Unlike other algorithms, here I used Block matching method instead of pixel based comparison of each frames.

This reduces the computational time and mismatching of the pixel was also avoided. Thus noises can be reduced. And also rather than change detection mask, background difference mask is considered. This also improves the quality of the video object detection.

Figure 2: Traffic video

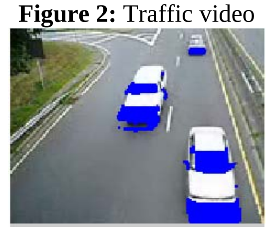

Figure 3: Moving object detected

8. Conclusions

This paper, presents an approach to recognize moving object in a video sequence. A combination of spatio-temporal spatial segmentation and temporal segmentation was used to detect the moving object. Instead of pixel based frame difference, I here considered block matching for the temporal segmentation process. This reduces the computational time and mismatching of the pixel was also avoided. Thus noises can be reduced. So I supposed to get a better result for the new proposed scheme. That is we can detect the moving objects in the video sequence more accurately.

References

[1] Krishna Kant Singh, Akansha Singh- A study of image segmentation algorithms for different types of images. IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010

[2] A. Papoulis and S. U. Pillai, Probability, Random Variables and Stochastic Process. New York: McGraw-Hill, 2002.

[3] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern., vol. 9, no. 1, pp. 62–66, Jan.1979.

[4] A. Ghosh, N. R. Pal, and S. K. Pal, “Image segmentation using a neural network,” Biol. Cybern., vol. 66, no. 2, pp. 151–158, 1991.

[5] A. Veeraraghavan, A. K. Roy-Chowdhury, and R. Chellappa, “Matching shape sequences in video with applications in human movement analysis,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 27, no. 12, pp. 1896–1909, Dec. 2005.

[6] N. Anjum and A. Cavallaro, “Multifeature object trajectory clustering for video analysis,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 11, pp. 1555–1564, Nov. 2008.

[7] G. H. Sasaki and B. Hajek, “The time complexity of maximum matching by simulated annealing,” J. ACM, vol. 35, no. 2, pp. 387–403, 1988.

[8] A. L. Bovic, Image and Video Processing. New York: Academic Press, 2000.

[9] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 6, no. 6, pp. 721–741, Nov

Author Profile

Hima E P received the B. Tech degrees in Information Technology from Co-operative Institute of Technology in 2011. She is currently doing her M. Tech degree from KMCT College of engineering.