

Moving Objects Detection from Video Sequences Using Fuzzy Edge Incorporated Markov Random Field Modeling and Local Histogram Matching

Badri Narayan Subudhi and Ashish Ghosh

Machine Intelligence Unit,
Indian Statistical Institute, Kolkata, India-700108
subudhi.badri@gmail.com, ash@isical.ac.in

Abstract. In this article, we put forward a novel region matching based motion estimation scheme to detect objects with accurate boundaries from video sequences. We have proposed a fuzzy edge incorporated Markov Random Field (MRF) model based spatial segmentation scheme that is able even to identify the blurred boundaries of objects in a scene. Expectation Maximization (EM) algorithm is used to estimate the MRF model parameters. To reduce the complexity of searching, a new scheme is proposed to get a rough knowledge of maximum possible shift of objects from one frame to another by finding the amount of shift in positions of the centroid. Moving objects in the scene are detected by the proposed χ^2 -test based local histogram matching. It is noticed that the proposed scheme provides better results with less object background misclassification as compared to optical flow and label fusion based techniques.

1 Introduction

Detection of moving objects in a given video has becoming an important research area in both computer vision and pattern recognition community because of its wide application in video surveillance, driving assistance system, robotics etc. It is very difficult to detect a moving object from a video captured with moving camera. To solve such problems one can adhere to the approaches like optical flow or motion estimation scheme [7]. However this scheme is unable to give good results for objects with complex motions and dynamic background. A combination of spatial segmentation and motion prediction/estimation is proved to be a better approach towards this.

An early work for object detection by region growing scheme is suggested by Deng and Manjunath in [2], where the spatial segmentation of an image frame is obtained by color quantization followed by an edge preserving region growing scheme. For temporal segmentation, the regions corresponding to objects are matched in the temporal direction by computing the motion vectors of the object regions in the target frame. Gray level values of pixels in a video with high uncertainty and high ambiguity make it difficult to detect moving objects with affordable accuracy by non-statistical spatial segmentation schemes. Hence, it requires some kind of stochastic method to model the attributes of an image

frame so that a better segmentation result can be obtained. Markov Random Field (MRF) model, in this context, proved to be a better framework. A robust work on MRF based object detection is demonstrated by Kuo *et al.* in [5], where a combination of temporal and spatial constraints of the image frames are used to obtain the moving object location from one frame to another of a video. Recently, Jodoin *et al.* [4] have proposed a robust moving object detection and tracking scheme for both fixed and moving camera captured video sequences, where MRF is used for label fields fusion. The label fields include two data: quickly estimated labels (i.e., a rough estimate of the location of object) and the spatial region map obtained by spatial segmentation with MRF model.

In this article, we propose a moving object detection scheme, that is able to detect moving objects with accurate object boundary from videos captured by moving camera. The technique uses a region based motion estimation scheme. Here we have proposed a new kind of MRF-MAP framework, where fuzzy edge strength at each pixel is incorporated in the MRF modeling. The scheme is able to preserve the object boundary. The spatial segmentation problem is solved using the Maximum a posteriori probability (MAP) estimation principle. In region based motion estimation scheme, to reduce the complexity of searching, a rough knowledge of maximum possible shift in object from one frame to another is obtained by calculating the amount of shift in the centroid of the object from one frame to another. Moving objects in the target frame is detected by χ^2 -test based local histogram matching. It is observed that the proposed scheme provide better results with less object background misclassification as compared to existing techniques.

2 Spatial Segmentation Using Proposed Fuzzy Edge Incorporated MRF Model

In this work it is assumed that the observed video sequence y is a 3-D volume consisting of spatio-temporal image frames. y_t represents a video image frame at time t and hence is a spatial entity of size $M \times N$. Each pixel in y_t is assumed as a site s denoted by y_{st} . Let us assume that X_t represents the MRF from which x_t is a realization. Here $X_t = X_t(i, j)$ is discrete valued and can take values from $Q = \{q_1, q_2, \dots, q_m\}$. It is assumed here that due to the presence of noise (i.i.d) we are not able to observe X_t , and we observe a noisy version of X_t as Y_t .

In spatial domain, X_t represents the MRF model of x_t and using Hammersley Clifford's theorem the prior probability can be expressed as Gibbs's distribution with $P(X_t) = \frac{1}{z} e^{-\frac{\bar{U}(x_t)}{T}}$, where z is the partition function expressed as $z = \sum_{x_t} e^{-\frac{\bar{U}(x_t)}{T}}$, $\bar{U}(x_t)$ is the energy function, a function of clique potentials $V_c(x_t)$. According to Pott's model the clique potential function $V_c(x_t) = -\alpha$ if all labels in possible set of cliques (C) are equal, otherwise $V_c(x_t) = +\alpha$. Equal penalization of all the boundary pixels results in a greater penalty to weak edge pixels and a lesser penalty to the strong edge pixels. This results in improper identification of boundary pixels in low resolution or blurred images. To reduce this effects one can adhere to the concept of incorporating local statistics based kernel function in MRF model as $\bar{U}(x_t) = \sum_n e^{-(x_t \otimes h)}$. This defines the energy

function in MRF as a function of x_t convolved with some local statistic based edge sensitive kernel h .

Generally the change in gray level between the successive regions in an image frame is very common and edge detection techniques are found to be effective only for significant contrast. Hence the inclusion of a deterministic edge kernel in MRF modeling is not expected to yield a satisfactory solution. It justifies to apply the concept of fuzzy set based edge kernel [8] rather than a deterministic edge kernel. The operation of fuzzy edge kernel at a particular site (with a set of neighboring pixels) makes the energy function as $\bar{U}(x_{st}) = \sum_{\eta_s} e^{-\frac{(x_{st} \otimes h)}{F_c}}$, where $(x_{st} \otimes h) = \frac{1}{(1 + \frac{|x_{st} - \bar{x}_{rt}|}{F_d})^{F_c}}$. Here F_e and F_d are two positive constants and are termed as the exponential and denominational fuzzifiers, respectively. The constant F_c is an MRF convergence parameter. There is no closed form solution for the estimation of parameters F_c , F_e and F_d and hence are fixed manually. Here $\hat{x}_{rt}$ represents $\max_{r \in \eta_s} \{x_{rt}\}$ or $\min_{r \in \eta_s} \{x_{rt}\}$. Hence we may write the prior probability $P(X_t)$ as $P(X_t) = e^{-\bar{U}(x_t)} = \frac{1}{z} \sum_{\eta_s} e^{-\frac{(x_t \otimes h)}{F_c}}$. The corresponding expression for MRF modeling with deterministic edge kernel can be obtained by replacing the fuzzy edge kernel with a deterministic edge kernel and termed as deterministic edge incorporated MRF modeling (DEMRF).

Here the segmentation problem is considered to be a process of determining a realization x_t that has given rise to the actual image frame y_t . One way to estimate $\hat{x}_t$ is based on the statistical MAP estimation by Bayes' theorem [6] as $\hat{x}_t = \arg \max_{x_t} P(Y_t = y_t | X_t = x_t) \theta P(X_t = x_t)$, where θ is the parameter vector associated with estimation of x_t . The likelihood function $P(Y_t = y_t | X_t = x_t, \theta)$ can be expressed as a realization of Gaussian noise $N(x_t, \sigma)$. Hence including the prior probability and the likelihood function we can rewrite the a posterior probability of MRF as

$$\hat{x}_t = \arg \max_{x_t} \left\{ A - \left[\frac{\|y_t - x_t\|^2}{2\sigma^2} \right] - \left[\sum_{\eta_s} \frac{(x_t \otimes h)}{F_c} \right] \right\}, \quad (1)$$

where $A = -\frac{1}{2} \log((2\pi)^3 \sigma^6)$. $\hat{x}_t$ in eq (1) is the MAP estimate of x_t . We have considered a combination of both simulated annealing (SA) and iterated conditional mode (ICM) algorithm [9] for estimating the MAP of each incoming image frame and the parameter $\theta = \{\sigma^2\}$ is estimated recursively by EM algorithm [6].

3 Region Matching Based Motion Estimation

In the initial phase of the proposed object detection framework, the candidate frame (i.e., the frame in which object position is already known) and the target frame (i.e., the frame in which object is required to be detected) both are spatially segmented by the proposed fuzzy edge incorporated MRF modeling approach. The object detection task in different frames of the given video is accomplished by matching the gray level distributions of those regions corresponding to each moving object in the candidate frame to a region in the target

frame by χ^2 -test based histogram matching. To get an effective object detection, we propose an algorithm to find the rough estimation of maximum possible shift in object from one frame to another by calculating the amount of shift in the centroid of the object from one frame to another.

Calculation of Shift in Centroid and Local Histogram Matching: Let us consider two image frames at t^{th} and $(t+d)^{th}$ instant of times, where the t^{th} frame is the candidate frame and $(t+d)^{th}$ frame is the target frame. Let us assume that there are d numbers of moving objects in the scene, and their positions in the candidate frame are known. From one frame to another as the object in the scene move, there will be a shift in the centroid of the object. Considering the relative camera movements in the scene as a multiplicative factor we can calculate the amount of shift of object in the scene as a function of distance. In the $(t+d)^{th}$ frame, the maximum possible shift of object in x-direction can be given as $shift_{(x,(t+d)b)} = const * \max\{|Dist_{(t+d),b(i)} - Dist_{t,b(i)}|_{b=1,2,...,d}\}$. Similarly, in y-direction it can be given as $shift_{(y,(t+d)b)} = const * \max\{|Dist_{(t+d),b(j)} - Dist_{t,b(j)}|_{b=1,2,...,d}\}$. The $Dist$ function represents distance of b^{th} object centroid in the candidate frame to the target frame image centroid. The $const$ represents a +ve constant that resolves the camera movements and scaling of the object in the scene.

For matching each region corresponding to a moving object in the candidate frame to a region within the search space in the target frame, a search is made up to $shift_{(x,(t+d)b)}$ in the +ve and -ve directions of the x-axis and $shift_{(y,(t+d)b)}$ in the +ve and -ve directions of the y-axis. Each time, we search a new region in the target frame, the histogram corresponding to the new region is matched with a region corresponding to the object in the candidate frame by χ^2 -test [1]. By χ^2 -test based measure, the distribution/histogram of two regions can be compared by the formula,

$$\chi^2((h^{(b_q, y_t)}), (h^{(g, y_{(t+d)})})) = \sum_{l=0}^{l=M-1} \frac{|(h^{(b_q, y_t)}) - (h^{(g, y_{(t+d)})})|}{(h^{(b_q, y_t)}) + (h^{(g, y_{(t+d)})})}.$$

The term $h^{(b_q, y_t)}$ represents the histogram corresponding to a region in b^{th} object of the candidate frame. The term $h^{(g, y_{(t+d)})}$ represents the histogram corresponding to a region within the search space (as obtained by centroid shifting scheme) in the target frame. Here g represents any region in the target frame within the obtained search space. M is the number of possible gray level bins in the histogram.

4 Results and Discussion

The proposed scheme is tested on several video sequences, however for page constraint we have provided the results on one test video sequence. The results obtained by the proposed fuzzy edge map incorporated MRF modeling for spatial segmentation is compared with those of the spatial segmentation by conventional MRF modeling [4] and MRF modeling with deterministic edge kernel (DEMRF). Similarly, the moving object detection results obtained by the proposed scheme are compared with those obtained by optical flow [7] and label fusion [4] schemes.

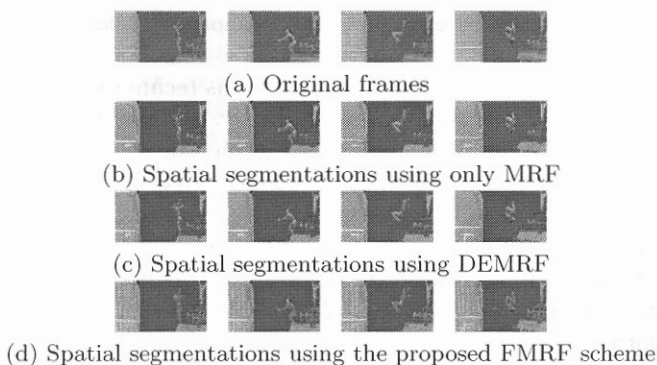


Fig. 1. Spatial segmentation results for Diving video sequence

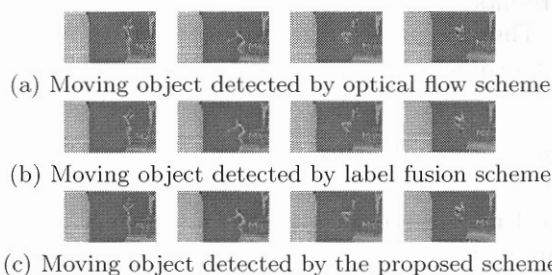


Fig. 2. Moving object detection for Diving video sequence

The video considered for our experiments is *Diving* video sequence having a single moving object, i.e. a gymnast. The considered frames of the sequence are shown in Fig. 1 (a). The spatial segmentation results obtained by the conventional MRF model and DEMRF are shown in Figs. 1 (b) and (c), respectively. It is observed from these results that the output obtained by these approaches are over-segmented. This is due to the fact that, the gray-level variation of one region to another of the image frames are very less. The boundary of the object in the scene are blurred and hence are indistinguishable. The results obtained by the proposed fuzzy edge incorporated MRF modeling (FMRF) scheme is shown in Fig. 1 (d). For evaluating the accuracy, we have used the pixel by pixel comparison of the manually constructed *ground-truth* images with the obtained spatial segmentation results. This measure is also called *number of misclassified pixels* (NMP). It is found that for the considered sequence, average NMP measure obtained for the conventional MRF scheme and the deterministic edge incorporated MRF are 2179.25 and 1777, respectively. The average NMP measure obtained for the proposed scheme is found to be 1114, much better than the existing methods. The time taken by the proposed scheme is 55 seconds.

The location of the moving object obtained for different frames of this sequence using optical flow based scheme are shown in Fig. 2 (a). It is found from

these results that the diving gymnast was not properly identified. Similarly, the moving object location obtained by the label fusion scheme are shown in Fig. 2 (b). It is observed from these results that this technique is also not able to identify the different parts of the gymnast correctly. However the results obtained by the proposed scheme (as shown in Fig. 2 (c)) has correctly detected the gymnast in different frames of the considered sequence. To evaluate the performance of the detected moving object locations, we have considered the manually constructed ground-truth based performance evaluation scheme of Erdem *et al.* [3]. This scheme utilizes four different measures: *misclassification penalty* (Mi.P), *motion penalty* (M.P), *shape penalty* (S.P), and *combined penalty* (C.P) to evaluate the performance of the moving objects locations. It may be noted that all these measures should be low for better detection of moving objects. These measures obtained for optical flow scheme are 0.5, 0.4, 0.4 and 0.45, respectively. Similarly these measure obtained for Label fusion scheme are 0.51, 0.38, 0.31 and 0.39, respectively. The results obtained for the proposed scheme are 0.42, 0.22, 0.22 and 0.29, respectively. Thus the proposed scheme is seen to provide better results as compared to the optical flow and label fusion schemes.

5 Conclusion

A new region based motion estimation scheme to detect moving objects from given video sequences has been formulated. We have proposed a spatial segmentation technique by incorporating the fuzzy edge strength of pixels in MRF modeling to preserve the object boundary. In this approach initially a rough estimate of object locations in the scene are obtained by finding the amount of shift in the centroid from one frame to another. The location of moving objects in a scene is obtained by χ^2 -test based local histogram comparison. The proposed scheme is found to be better as compared to the existing moving object detection schemes. The proposed scheme does not yield good results, with occlusion/dis-occlusion condition. In our future work we will try to solve the above problem.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J.(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
2. Deng, Y., Manjunath, B.S.: Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 23(8), 800–810 (2001)
3. Erdem, C.E., Sankur, B., Tekalp, A.M.: Performance measures for video object segmentation and tracking. *IEEE Transactions on Image Processing* 13(7), 937–951 (2004)
4. Jodoin, P.M., Mignotte, M., Rosenberger, C.: Segmentation framework based on label field fusion. *IEEE Transactions on Image Processing* 16(10), 2535–2550 (2007)

5. Kuo, C.M., Hsieh, C.H., Huang, Y.R.: Automatic extraction of moving objects for head-shoulder video sequence. *Journal of Visual Communication and Image Representation* 16(1), 68–92 (2005)
6. Li, S.Z.: *Markov Random Field Modeling in Image Analysis*. Springer, Japan (2001)
7. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *Proceedings of Imaging understanding workshop*, pp. 121–130 (1981)
8. Pal, S.K., King, R.A.: On edge detection of x-ray images using fuzzy sets. *IEEE Transactions on Pattern Analysis Machine Intelligence* 5(1), 69–77 (1983)
9. Subudhi, B.N., Nanda, P.K., Ghosh, A.: Moving object detection using MRF model and entropy based adaptive thresholding. In: *Proceedings of IEEE 2nd International Conference on Human Computer Interaction*, vol. 1, pp. 155–161. Springer, Heidelberg (2010)