Moving Object Detection Using MRF Model and Entropy based Adaptive Thresholding

Badri Narayan Subudhi, Pradipta Kumar Nanda and Ashish Ghosh

Abstract. In this article, we propose an algorithm for moving object detection which includes two segmentation schemes: spatial segmentation and temporal segmentation. For spatial segmentation we have used a compound Markov Random Field (MRF) Model for attribute modeling in the temporal direction and the corresponding problem is formulated using maximum a posteriori probability (MAP) estimation principle. For temporal segmentation we propose a label frame difference based change detection mask (CDM). The difference image threshold value is obtained by the proposed entropy based adaptive windowing scheme. A combination of both spatial and temporal segmentations detects the moving objects. It is observed that entropy based adaptive windowing scheme gives better results towards moving object detection with less effect of silhouette than using the non window based global thresholding approach.

Keywords: Object detection, MAP estimation, Modeling, Simulated annealing, Entropy, Threshold, Gaussian distribution, Image segmentation

1 Introduction

Detection of moving objects from a video scene is a challenging task in Video Processing and Computer Vision [1]. Based on the movements of objects and background, video sequences can be categorized into two types, such as moving objects with moving background and moving object with fixed background. In the later case moving object detection can be accomplished by motion detection/temporal segmentation approach alone. But, it does not yield good result in absence of reference frame or for the case where tangible movement of objects is not there. A combination of both temporal segmentation and spatial segmentation proved to be a better approach towards this [1]. Moving objects detection highly depends on the robustness of the spatial segmentation. Markov Random Field (MRF) model [2], in this context, is proved [3] to be better. An early work on MRF based object detection scheme was proposed by Hwang et al. [3]. In this approach the spatial segmentation was obtained with attribute modeling by MRF and Distributed genetic Algorithm (DGA) was used for MAP estimation. Similarly the Temporal segmentation was obtained by direct combination of video object plane (VOP) of the previous frame with the
current frames change detection mask (CDM). The objects of the previous frame were assumed to be present in the current frame also. This assumption created difficulty in exact detection of moving objects and gave rise to an effect of silhouette. The effect of silhouette was less in the recently proposed algorithm of Kim et al. [4]. They had used an evolutionary probability to update the crossover and mutation rate through evolution in DGA.

In order to reduce the effect of noise and illumination variation in the object detection, Su et al. [5] had proposed an adaptive thresholding approach of change detection. In this approach, each difference image was divided into a number of blocks and is tested for presence of region of change (ROC) with an ROC scatter estimation algorithm. The threshold values all marked block containing ROC was obtained and were averaged to obtain a global threshold.

In this article, for spatial segmentation, we have used an edge-based compound MRF model for attribute modeling of a given video image frames [6] followed by the MAP estimation with a hybrid algorithm (hybrid of both simulated annealing (SA) and iterated conditional mode (ICM)). The compound MRF uses spatial distribution of color in the current frame, color coherence in the temporal direction and edge maps in the temporal direction. The difference images obtained from the given video frames are largely affected by illumination variation and noise that propagates in the form of silhouette to the VOP. Hence, we propose an adaptive temporal segmentation scheme that reduces the effect of noise. Instead of segmenting the whole image at a time by a single threshold, we partition the input image into different windows/blocks and segment the objects in each of these windows. Then we combine the segmented objects from each window. The window size is determined by the entropy content of the considered window. It is observed that the proposed entropy based adaptive window scheme gives better results than that of non window based global thresholding approach [7], where a large effect of silhouette exists. The spatial segmentation is combined with adaptive thresholding based temporal segmentation to construct the VOPs and hence moving object detection.

The organization of this article is as follows. Section 2 represents the moving object detection scheme using spatio-temporal spatial segmentation and proposed temporal segmentation scheme. Section 3 represents simulation and result analysis. Conclusion is presented in Section 4.

2 Moving Object Detection

In this scheme moving objects are determined as follows. In a given image frame, the spatial segmentation is obtained by modeling the attributes of the image frames by an edge-based compound MRF Model followed by the MAP estimation with a hybrid algorithm [6]. The difference image obtained from the given image frames is thresholded by the proposed entropy based adaptive window scheme. The thresholded image is fused with the labels obtained from the spatial segmentation result to obtain the final temporal segmentation. Subsequently, the pixels corresponding to the foreground part of the temporal segmentation is used to display the VOP.
2.1 Spatial Segmentation: Spatio-temporal image Modeling and MAP Estimation

Here it is assumed that the observed video sequence $y$ is a 3-D volume consisting of spatio-temporal image frames. $y$ represents the video image frame at time $t$. Each pixel in $y_t$ is a site $s$ denoted by $y_{st}$. Let $Y_t$ represent a random field and $y_t$ be a realization of it at time $t$. Thus, $y_{st}$ denotes a spatio-temporal co-ordinate of the grid $(s, t)$. Let $x$ denote the segmentation of video sequence $y$ and $x_t$ the segmented version of $y_t$. Let us assume that $X_t$ represent the MRF from which $x_t$ is a realization. Similarly the pixels in the temporal direction are also modeled as MRFs. We have considered the second order MRF modeling both in spatial and in temporal directions. In order to preserve the edge features, another MRF model is considered with the linefield/edge map of the current frame $x_t$ and the linefields/edgemaps of $x_{t-1}$ and $x_{t-2}$.

In spatial domain, the prior probability can be expressed as Gibb’s distribution $p(x_t) \propto \exp(-U(x_t))$, where $z$ is the partition function expressed as $z = \sum \exp(-U(X))$. $U(X)$ is the energy function (a function of clique potential). We have considered the clique potential in spatial direction as $V_c(x_t) = -\alpha$ if all labels in possible set of cliques ($C$) are equal, otherwise $V_c(x_t) = \alpha$. Analogously in the temporal direction, $V_{tec}(x_t) = -\beta$ if all labels in $C$ are equal, otherwise $V_{tec}(x_t) = \beta$ and for the edgemap in the temporal direction as $V_{tec}(x_t) = -\gamma$ if all labels in $C$ are equal, otherwise $V_{tec}(x_t) = \gamma$. We have used the additional feature in the temporal direction and the whole model is referred as edgebased model.

In our a priori image modeling the clique potential function is the combination of the above three terms. Hence, the prior energy function is of the following form

$$U(X_t) = \sum_{c \in C} V_c(x_t) + \sum_{c \in C} V_{tec}(x_t) + \sum_{c \in C} V_{tec}(x_t).$$

(1)

The observed image sequence is assumed to be a degraded version of the actual image sequence $x$. The degradation process is assumed to be Gaussian. Thus, the label field $x_t$ can be estimated from the observed random field $Y_t$. The label field is estimated by maximizing the following posterior probability distribution.

$$\hat{x}_t = \arg \max_{x_t} \frac{P(Y_t = y_t | X_t = x_t) p(X_t = x_t)}{P(Y_t = y_t)},$$

(2)

where $\hat{x}_t$ denotes the estimated labels. The prior probability $P(Y_t = y_t)$ is constant and can be discarded. Assuming decorrelation of the three RGB planes for the color image and the variance to be the same among each plane, the likelihood function $P(Y_t = y_t | X_t = x_t)$ can be expressed as

$$P(N = y_t - x_t | X_t, \theta) = \frac{1}{\sqrt{(2\pi)^3 \sigma^3}} e^{-\frac{1}{2\sigma^2} (y_t - x_t)^2},$$

(3)
In (3), variance \( \sigma^2 \) corresponds to the Gaussian degradation. Here \( n \) is a realization of the Gaussian noise \( N(\mu, \sigma) \). Now using eqs (1) and (3) in eq (2) we may obtain

\[
x_n = \arg \min_x \left( \frac{1}{2\sigma^2} \sum_{r \in \Gamma} v(x_n) + \sum_{i \in \partial} v(i) + v_{\text{err}}(x_n) \right).
\]

\( \hat{x} \) is the MAP estimate and is obtained by a hybrid algorithm (hybrid of SA and ICM algorithm)[6].

### 2.2 Temporal Segmentation

If a global thresholding algorithm is applied in a difference image, affected by noise, illumination variation or shading: (i) few pixels that actually correspond to the background in an image frame are identified as changed pixels, whereas these are actually not, (ii) it also happens that a pixel in the difference image that correspond to actual change region and lies in the lower range of the histogram may be identified as an unchanged pixel. An adaptive thresholding approach can be used to overcome these problems. However the choice of window size is an issue.

In this regards we propose an entropy based adaptive window selection scheme to determine the block/window size. Here the threshold value for a particular window is obtained by Otsu’s thresholding [7] scheme. To enhance the segmentation results, the results thus obtained from CDM are verified and compensated by considering the information of the pixels belonging to objects in the previous frame. This is represented as

\[
R = \{ r_s | 0 \leq s \leq (M - 1) \times (N - 1) \}, \tag{5}
\]

where \( R \) is a matrix having the same size of the frame, \( s \) is the element number in the matrix and \( r_s \) is the value of the VOP at location \( s \). If a pixel is found to have \( r_s = 1 \), it is a part of the moving object of the previous frame; otherwise it belongs to the background of the previous frame. Based on this information, CDM is modified as follows: if it belongs to a moving object part in the previous frame and its label obtained by spatio-temporal segmentation is the same as one of the corresponding pixels in the previous frame, the pixel is marked as the foreground area in the current frame else as a background. The modified CDM thus represents the temporal segmentation result.

**Entropy Based Window Growing**: The basic notion of window growing approach is to fix the window size primarily focusing on the information measure of the image at different scales. In other words, fixing the size of the window depends on the entropy of the chosen window. In this approach an arbitrarily small window (here the window size \( w \) is chosen as \( 5 \times 5 \)) is considered initially and the entropy of the window is computed from the gray level distribution of the window and is denoted by \( H_w \)

\[
H_w = \sum_{i=1}^{G} p_i \log \left( \frac{1}{p_i} \right) \tag{6}
\]

where \( p_i \) is the probability of occurrence of \( i \)th gray level and \( G \) is the maximum
gray level. If the entropy of the window is comparable to some fraction of the entropy of the whole image, (represented as $Th$), that window is chosen for segmentation (by Otsu’s thresholding [7]); otherwise the window will be incremented by $\Delta w$ (here $\Delta w$ is considered 2) and the condition will be tested again. The window will be fixed if the total image is exhausted. The final thresholded image is obtained by taking union of each considered thresholded windows.

3 Simulations and Result Discussion

In order to test the effectiveness of the proposed approach, we have tested our algorithm in Canada traffic video sequence as presented in Fig. 1. This video contains three objects such as a black car, a white car and a person moving at different speeds. Fig. 1(a) shows the original 3rd, 4th, 5th and 6th frames of Canada traffic video sequence. The corresponding spatial segmentation results obtained by edgebased model approach are shown in Fig. 1(b). The MRF model Parameters used for Canada traffic video sequence are $\alpha = 0.01$, $\beta = 0.009$, $\gamma = 0.007$ and $\sigma = 3.0$. Here the MRF model parameters are determined on trial and error basis. JSEG based spatial segmentation of these frames are displayed in Fig. 1(c). It is observed from these results that the region containing the black car (left side of the image) and the person in the lawn are merged into background hence very difficult to identify. It is to be noted that the edge based spatial segmentation scheme could segment all the moving parts and the static parts while preserving the boundary accurately. As seen from Fig. 1(c), use of global thresholding in temporal segmentation could not detect the black car as well as the man properly and hence the effect of silhouette is observed. As observed from Fig. 2(f), use of adaptive threshold could detect all the three moving objects (black car, white car and the person) properly. In order to measure the performance quantitatively, ground truth images have been created manually. The misclassification errors obtained for different frames are: 4th frame: 82 pixels, 5th frame: 91 pixels and 6th frame: 80 pixels. Correspondingly Otsu’s method produced 151, 131 and 124 pixels of error.

4 Conclusions

In this article, a problem of moving object detection is addressed. A compound MRF model is used here to model both spatial and temporal attributes of the video image frames. Corresponding MAP estimate is obtained by a hybrid algorithm that converges fast. For temporal segmentation we have used a label frame difference as opposed to an original image frame difference. The threshold value for the difference image is determined using an adaptive thresholding algorithm whose window size is chosen by the entropy measure over the window. It is observed that this approach gives better results compared to original frame difference CDM followed by Otsu’s thresholding approach [7], as
the effect of silhouette is less.

References


Fig. 1. VOP generation of Canada traffic video sequences