# BACKGROUND SUBTRACTION USING BELIEF PROPAGATION

Hee-il Hahn
*Dept. Information and Communications Eng., Hankuk University of Foreign Studies, Yongin, Korea*

Keywords: Background subtraction, Pixel-based background modelling, Visual surveillance, Markov random fields, Belief propagation.

Abstract: It is challenging to detect foreground objects when background includes an illumination variation, shadow or structural variation due to their motion. Basically pixel-based background models suffer from statistical randomness of each pixel. This paper proposes an algorithm that incorporates Markov random field(MRF) model into pixel-based background modelling to achieve more accurate foreground detection. Under the assumptions the distance between the pixel on the input image and the corresponding background model and the difference between the scene estimates of the spatio-temporally neighboring pixels are exponentially distributed, a recursive approach for estimating the MRF regularizing parameters is proposed. The proposed method alternates between estimating the parameters with the intermediate foreground detection results and detecting the foreground with the estimated parameters, after computing them with the detection results of the pixel-based background subtraction. Extensive experiment is conducted with several videos recorded both indoors and outdoors to compare the proposed method with the codebook-based algorithm.

## 1 INTRODUCTION

Computer vision systems such as visual surveillance, object tracking need to separate the moving objects from the scene background. Background subtraction in the field of view of stationary video camera is a common approach for detecting foregrounds from the dynamic backgrounds. Usually background subtraction employs pixel-based background model. Its simplest model assumes a pixel can be modelled with statistical informations such as mean and variance estimated from the corresponding pixel location of a sequence of video frames. This method tries to detect the foreground by thresholding the intensity or color difference between the current frame and the background model. However it is very sensitive to the selection of threshold and rarely deals with the dynamics of backgrounds, like illumination variations or the local motion of the background objects, *e.g.* waving trees. Their dynamics causes the pixel intensity values to vary significantly with time. Many authors proposed several promising schemes to model such variations. Among them are the generalized mixture of Gaussians (Stauffer and Grimsom, 1999), nonparametric kernel model (Elgammal, et al., 2002), or codebook model (Kim, et al., 2005), etc. Stauffer and Grimsom model the pixel intensity with a mixture of 3 to 5 Gaussian distributions and use the EM algorithm for adaptation of the mixture model. Elgammal, et al. estimate the density function of each pixel nonparametrically using a kernel function. When the Gaussian kernel function is adopted it can be viewed as a generalization of the Gaussian mixture model. Kim, et al. adopt a codebook quantization scheme to construct a background model from long observation sequences. Each background pixel has a codebook composed of group of codewords. Although a single codeword may be enough to model static background pixel, mixed background pixel can be modelled by multiple codewords whose number depends on the dynamics of the pixel. All the above approaches are similar in that they handle the complex backgrounds by modelling a pixel with multi-modal distributions. However, Pixel-based algorithms like the above approaches basically assume the statistics of each pixel are independent although they are highly correlated with the neighboring ones. Some researchers employ the block-based models or Markov random field techniques to improve the pixel-based algorithms. MRF-based methods usually

Hahn H..
BACKGROUND SUBTRACTION USING BELIEF PROPAGATION.
DOI: 10.5220/0003444102810286
In *Proceedings of the 8th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2011), pages 281-286
ISBN: 978-989-8425-75-1

exploit the spatial and temporal dependencies of the pixels by developing MRF models for background subtraction. MRF assumes that each variate corresponding to its pixel location is connected to its four or eight nearest neighbours. MRF needs cost functions which are related with the compatibility functions between the scene variable and the corresponding pixel value. Basically any background model can be used to define the cost functions. This paper chooses a codebook-based background model for the cost functions. Almost all MRF-based background models select the fixed values for all MRF parameters. For example, (Migdal, et al., 2005) assigns the constant energy potentials for all the spatial, posterior and temporal cliques and (Wu, et al., 2010) assumes all compatibility functions are exponentially distributed with constant parameters. (McHugh, et al., 2009) models the background subtraction as a binary hypothesis test and determines the detection threshold by means of Ising model. (Xu, et al., 2008) recovers the background image from a sequence of images containing moving foreground objects. A loopy belief propagation is employed for background estimation.

A loopy belief propagation is also adopted in this paper. However its roles are quite different in that it decides whether an image pixel belongs to background or foreground in this paper, while Xu, et al. use it to indicate from which frame the pixel should be selected.

This paper makes major contributions that exploits both the spatial and temporal dependencies by developing MRF models for background subtraction and proposes a recursive approach for estimating the MRF regularizing parameters.

## 2 MRF-BASED FOREGROUND DETECTION

Let $X=\{x_i\}$ denote a set of binary random variable, where $i$ represents a pixel location. A state space is assumed, say $\Lambda=\{0,1\}$, so that $x_i \in \Lambda$ for all $i$. Let $\Omega$ be the set of all possible configurations:

$$\Omega=\{\omega=(x_1,x_2,\cdots,x_N): x_i \in \Lambda, 1\le i\le N\} \quad (1)$$

And a set of random variable $X$ is assumed to be a MRF. Then the probability $P(X=\omega)$ is a Gibbs distribution, depicted as:

$$P(X=\omega)=\frac{1}{Z}e^{-\frac{U(\omega)}{T}} \quad (2)$$

where $Z$ is a normalizing constant called the partition function, $T$ is a constant called the temperature and $U(\omega)$ is the energy function. The energy is a sum of clique potentials $V_c(\omega)$ over all possible cliques $c\in\mathbb{C}$, which is defined as

$$U(\omega)=\sum_{c\in\mathbb{C}}V_c(\omega)=\sum_i V_i(x_i)+\sum_{i,j}V_{i,j}(x_i,x_j) \quad (3)$$

For MRF-based background model, a superscript is added to the random variable $x_i$ so that $x_i$ is replaced with $x_i^t$, where $t$ represents a time index. The energy function $U(\omega)$ is extended in the following way, to include the time dependency as well as the spatial dependency.

$$\begin{aligned}U(\omega)&=\sum_{c\in\mathbb{C}}V_c(\omega)\\&=\sum_i V_i(x_i^t)+\sum_{i,j}V_{i,j}(x_i^t,x_j^t)+\sum_{i,j}V_{i,j}(x_i^t,x_j^{t-1})\end{aligned} \quad (4)$$

The scene variable $x_i^t$ is associated with the pixel value $y_i^t$ at time $t$ and pixel location $i$. That is, $x_i^t$ has a value of 0 when its corresponding pixel value $y_i^t$ comes from the background model and $x_i^t=1$ in case of foreground.

There is some statistical dependency between the pixel value $y_i^t$ at time $t$ and its corresponding decision result or scene variable $x_i^t$ at each pixel location $i$. A background pixel must come out from the background model, and so the potential $V_i(x_i^t)$ in (4) measures how the background pixel deviates from the background model, for the same case with the foreground pixel. Thus, $V_i(x_i^t)$ can be defined as:

$$V_i(x_i^t)=\begin{cases}\mu d(y_i^t) & y_i^t \in Background\\ \Gamma & y_i^t \in Foreground\end{cases} \quad (5)$$

where $\mu$ is the proportional constant and $\Gamma$ is the potential associated with the foreground pixel, which is optimally adjusted using the EM algorithm, as explained later in 2.2. And $d(y_i^t)$ can be obtained using any pixel-based background model. Since this paper employs the codebook model (Kim, et al., 2005), $d(y_i^t)$ is defined as a minimum distance between an input pixel $y_i^t$ and the centroids of the codeword $c_k$ belonging to the codebook $C_i$.

The node $i$ is arranged in a two-dimensional grid, and so its scene variable $x_i^t$ should be compatible with the nearby scene variables $x_j^t$. Let $\lambda$ be a

probability that $y_i^t$ will come out from the background model and $E_i\left(x_i^t\right)$ be the energy term corresponding to $V_i\left(x_i^t\right)$. Then $E_i\left(x_i^t\right)$ can be reduced to

$$E_i\left(x_i^t\right)=\lambda e^{-\mu d\left(y_i^t\right)}+\frac{1-\lambda}{M} \tag{6}$$

where $M=e^{\Gamma}$. Then (5) can be depicted as:

$$V_i\left(x_i^t\right)=-\log\left(\lambda e^{-\mu d\left(y_i^t\right)}+\frac{1-\lambda}{M}\right) \tag{7}$$

The potential $V_{i,j}\left(x_i^t,x_j^t\right)$ between $x_i^t$ and $x_j^t$ is defined so that it has a larger value when the variable $x_i^t$ is different from $x_j^t$, as follows:

$$V_{i,j}\left(x_i^t,x_j^t\right)=\nu\left|x_i^t-x_j^t\right| \tag{8}$$

Likewise, $V_{i,j}\left(x_i^t,x_j^{t-1}\right)$ is defined as

$$V_{i,j}\left(x_i^t,x_j^{t-1}\right)=\sigma\left|x_i^t-x_j^{t-1}\right| \tag{9}$$

where $\nu$ and $\sigma$ are the proportional constants. Then, (4) can be represented as:

$$\begin{aligned}U=&-\sum_i\log\left\{\lambda\varsigma e^{-\mu d\left(y_i^t\right)}+\frac{1-\lambda}{M}\right\}\\&+\sum_{(i,j)}\left(\nu\left|x_i^t-x_j^t\right|+\sigma\left|x_i^t-x_j^{t-1}\right|\right)\end{aligned} \tag{10}$$

The above equation can be further simplified by noting that a function of the form $-\log\left(ae^{-b|x|}+c\right)$ is tightly upper bounded by $\min\left(\beta|x|,\gamma\right)+\alpha$, where $\alpha=-\log(a+c)$, $\beta=\frac{ab}{a+c}$ and $\gamma=\log\left(\frac{a+c}{c}\right)$. Thus, minimizing (10) is equivalent to minimizing

$$\begin{aligned}U=&\sum_i\min\left(\kappa\left|d\left(y_i^t\right)\right|,\theta\right)+\\&+\sum_{(i,j)}\left(\nu\left|x_i^t-x_j^t\right|+\sigma\left|x_i^t-x_j^{t-1}\right|\right)\end{aligned} \tag{11}$$

where $\kappa=\frac{\lambda\mu}{\lambda+\frac{1-\lambda}{M}}$ and $\theta=\log\left(1+\frac{\lambda M}{1-\lambda}\right)$.

The belief propagation is adopted to solve the above equation (Yedidia, et. Al., 2002). Let $m_{ij}$ be the message that node $i$ sends to a neighboring node $j$ at time $t$. It is determined by the message update rules:

$$\begin{aligned}m_{ij}^t\left(x_j^t\right)=&\sum_i\min\left(\kappa\left|d\left(y_i^t\right)\right|,\theta\right)+\\&+\min\sum_{(i,j)}\left(\nu\left|x_i^t-x_j^t\right|+\sigma\left|x_i^t-x_j^{t-1}\right|\right)+\sum_{k\in N(i)/j}m_{ki}^t\left(x_i^t\right)\end{aligned} \tag{12}$$

And the belief $b_i\left(x_i^t\right)$ at a node $i$ is computed as

$$b_i\left(x_i^t\right)=\min\left(\kappa\left|d\left(y_i^t\right)\right|,\theta\right)+\sum_{k\in N(i)}m_{ki}\left(x_i^t\right) \tag{13}$$

where $N(i)$ denotes the nodes neighboring $i$. The scene variable $x_i^t$ is selected so that $b_i\left(x_i^t\right)$ should be minimized, namely

$$x_i^t=\begin{cases}1 & \mathrm{b}_i(0)>\mathrm{b}_i(1)\\0 & \mathrm{b}_i(0)\le\mathrm{b}_i(1)\end{cases} \tag{14}$$

## 2.1 Estimating $\nu$ and $\sigma$

The parameters $\nu$ and $\sigma$ are initialized using the detection results of the pixel-based background subtraction method. The energy term associated with the potential $V_{i,j}\left(x_i^t,x_j^t\right)$ corresponds to the joint probability, called the compatibility function, given as:

$$E_{i,j}\left(x_i^t,x_j^t\right)=e^{-\nu\left|x_i^t-x_j^t\right|} \tag{15}$$

So the probabilities corresponding to $x_i^t=x_j^t$ and $x_i^t\neq x_j^t$ are computed from the histogram of the detection results at time $t$, where $j$ is the neighbour of $i$. The parameter $\nu$ can be estimated as

$$\nu=-\log\left\{\frac{\sum_{(i,j)}h\left(x_i^t\neq x_j^t\right)}{\sum_{(i,j)}h\left(x_i^t=x_j^t\right)}\right\} \tag{16}$$

where $h(\cdot)$ is the histogram computed from the segmented image $X^t=\left\{x_i^t\right\}$.

Likewise, $\sigma$ can be obtained by

$$\sigma=-\log\left\{\frac{\sum_{(i,j)}P\left(x_i^t\neq x_j^{t-1}\right)}{\sum_{(i,j)}P\left(x_i^t=x_j^{t-1}\right)}\right\} \tag{17}$$

using $X^t$ and $X^{t-1}$.

## 2.2 Estimating $\mu$ and $\lambda$

The parameters $\mu$ and $\lambda$ are estimated using the expectation maximization algorithm. Let

$L = \max\left\{d\left(y_i^t\right)\right\}+1$ be the number of possible distance values of the pixels which come out from the background model. A random variable $\xi_i$ is assigned to each pixel $y_i^t$, indicating whether the pixel comes out from the background model. In other words, $\xi_i$ has a value of 0 when $y_i^t$ belongs to the background model, otherwise $\xi_i$ equals 1. Then the conditional probability of $\xi_i$ can be computed as

$$\begin{aligned}\rho_i &= P\left(\xi_i = 0 \middle/ d\left(y_i^t\right), \lambda, \mu\right) \\ &= \frac{\lambda e^{-\mu d\left(y_i^t\right)}}{\lambda e^{-\mu d\left(y_i^t\right)} + \frac{1-\lambda}{M}}\end{aligned} \tag{18}$$

Using the method proposed by (Zhang, Seits, 2007), the parameters $\mu$ and $\lambda$ are estimated by maximizing the expected log-probability $E_{\xi_i}\left[\log P\left(d\left(y_i^t\right), \xi_i \middle/ \lambda, \mu\right)\right]$, where $P\left(d\left(y_i^t\right), \xi_i \middle/ \lambda, \mu\right)$ is given as

$$\begin{aligned}P\left(d\left(y_i^t\right), \xi_i = 0 \middle/ \lambda, \mu\right) &= \lambda e^{-\mu d\left(y_i^t\right)} \\ P\left(d\left(y_i^t\right), \xi_i = 1 \middle/ \lambda, \mu\right) &= \frac{1-\lambda}{M}\end{aligned} \tag{19}$$

Using the above equations, $E_{\xi_i}\left[\log P\left(d\left(y_i^t\right), \xi_i \middle/ \lambda, \mu\right)\right]$ can be expressed as follows.

$$\begin{aligned}&E_{\xi_i}\left[\log P\left(d\left(y_i^t\right), \xi_i \middle/ \lambda, \mu\right)\right] \\ &= \sum_{y_i^t \in b} \rho_i \log P\left(d\left(y_i^t\right), \xi_i = 0 \middle/ \lambda, \mu\right) \\ &\quad + \sum_{y_i^t \in f} (1-\rho_i) \log P\left(d\left(y_i^t\right), \xi_i = 1 \middle/ \lambda, \mu\right)\end{aligned} \tag{20}$$

This equation can be reduced to be

$$\begin{aligned}&E_{\xi_i}\left[\log P\left(d\left(y_i^t\right), \xi_i \middle/ \lambda, \mu\right)\right] \\ &= \sum_{y_i^t \in B} \rho_i \left(\log(\lambda) - \mu\left|d\left(y_i^t\right)\right|\right) + \sum_{y_i^t \in F} (1-\rho_i) \log\frac{1-\lambda}{M}\end{aligned} \tag{21}$$

By setting the partial derivatives of the above equation with respect to $\lambda$ and $\mu$ to be zero, $\lambda$ is estimated as

$$\lambda = \frac{\sum_{y_i^t \in b} \rho_i}{\sum_{y_i^t \in B} \rho_i + \sum_{y_i^t \in F} (1-\rho_i)} \tag{22}$$

where $B$ and $F$ represent background and foreground, respectively. $\lambda$ actually can be approximated as the ratio of the number of pixels decided as background over the total number of pixels. And $\mu$ is the solution of the equation

$$\frac{1}{e^{\mu}-1} - \frac{L}{e^{\mu L}-1} = \frac{\sum_{y_i^t \in b} \rho_i \left|d\left(y_i^t\right)\right|}{\sum_{y_i^t \in b} \rho_i} \tag{23}$$

According to our experimentation results, $L$ is over 30, so that the second term of the left-hand side of (23) is negligible. Thus, the above equation can be solved explicitly as

$$\mu = \log\left(1 + \frac{1}{\chi}\right) \tag{24}$$

where $\chi$ is the right-hand side of (23).

The proposed method alternates between estimating the parameters with the intermediate foreground detection results and detecting the foreground with the estimated parameters, after computing them with the detection results of the codebook-based background subtraction.



## 3 EXPERIMENTAL RESULTS

The proposed method is tested with the real videos recorded indoors and outdoors, whose ground truths are manually segmented. Codebook algorithm (Kim, et. al., 2005) is selected as a pixel-based background model. Any postprocessing operations such as morphologies or connected component labelling are not used to demonstrate the effectiveness of the proposed method.

In the first sequence, illumination variation occurs according to the distance between the camera and the foreground object. Fig. 1 depicts the comparative detection results on the video recorded indoors. The third and fourth columns show the results of the codebook algorithm and the proposed method, respectively. The input frames show the lower limbs can rarely be identified from their background regions due to their slight color difference under the dark background, while the upper body of the object is very discriminative from the background. The proposed method detects the lower limbs more clearly than the codebook algorithm.

Fig. 2 shows the results on the video recorded outdoors. As can be seen from the input frame on the first column, the color of the lawn near the center region is very similar to that of her jacket. The codebook algorithm can not distinguish between them clearly and yields the streaks of false negatives

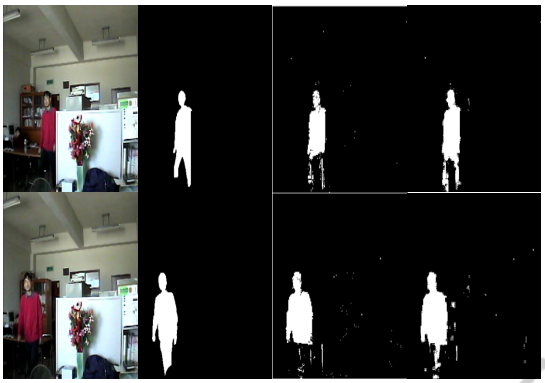


Figure 1: The comparative experimental results on the video recorded indoors. Column 1: original images. Column 2: ground-truths. Column 3: detection results of codebook-based algorithm. Column 4: detection results of our method.

near the middle of the detected foreground.

Basically the pixel-based algorithms can hardly distinguish the foreground objects from the background under the above situation. However, MRF can solve it by communicating with the adjacent pixels through the compatibility functions mentioned above. However, the proposed method misclassifies some background regions as foreground, which are revealed as small blobs scattered.

Fig. 3 shows the similarity test results to evaluate the performance of the proposed method quantitatively. The similarity (Chen, et. al., 2007) is defined as follows,

$$S(L_t, G_t) = \frac{(L_t \cap G_t)}{(L_t \cup G_t)} \quad (25)$$

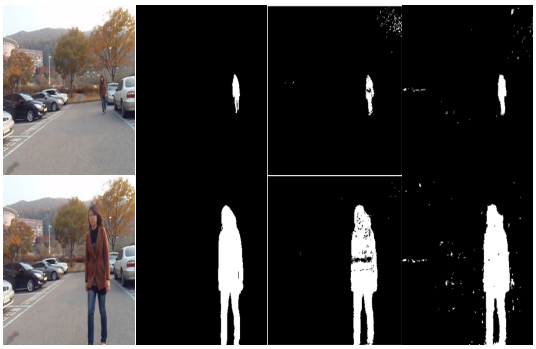


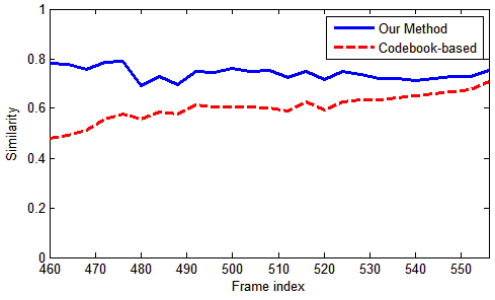
Figure 2: The comparative experimental results on the video recorded outdoors. Column 1: original images. Column 2: ground-truths. Column 3: detection results of codebook-based algorithm. Column 4: detection results of our method.

where $L_t$ and $G_t$ represent the detection result and the corresponding ground truth, respectively. The ground truths are manually segmented every 4 frame. The similarity value approaches 1 when the overlapped region between $L_t$ and $G_t$ increases. The proposed method shows the similarity value higher than that of the codebook algorithm at almost every frame. However at some frames of video recorded outdoors, the segmentation performance degrades slightly due to increase of false positives.

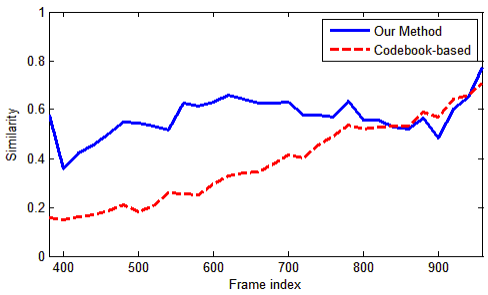
## 4 CONCLUSIONS

The algorithm that incorporates MRF model into the pixel-based background model is proposed. Basically almost all MRF-based background models select the fixed values for all MRF parameters. The proposed method shows the improved foreground detection by estimating all the parameters adaptively, instead of using the fixed parameters.

Extensive experiment conducted with videos recorded indoors and outdoors demonstrates the proposed MRF model effectively reduces the false negatives in detecting the foreground objects under complex background. However it is shown that the proposed method misclassifies some background regions as foreground slightly more, compared with the pixel-based segmentation algorithms. More efforts will be needed to reduce the number of such misclassifications without an appreciable degradation in classification speed.




(a)




(b)

Figure 3: The similarity curves for the codebook-based algorithm and the proposed method on the video recorded (a) indoors and (b) outdoors.

## REFERENCES


Stauffer, C., Grimson, W. E. L., 1999. Adaptive background mixture models for real-time tracking. *IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246-252.

Elgammal, A., Duraiswami, R., Harwood, D., Davis, L. S., 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proc. IEEE*, vol. 90, no. 7, pp. 1151-1163.

Kim, K., Chalidabhongse, T. H., Harwood, D., 2005. Real-time foreground-background segmentation using codebook model. *Elsevier Real-Time Imaging*, vol. 11, pp. 172-185.

Migdal, J., Grimson, W. E., 2005. Background subtraction using Markov thresholds. *Proceedings of the IEEE Workshop on Motion and Video Computing* (WACV/MOTION'05).

Wu, M., Peng, X., 2010. Spatio-temporal context for codebook-based dynamic background subtraction. *International Journal of Electronics and Communications*, pp. 739-747.

Chen, Y., Chen, C., Huang, C., Hung, Y., 2007. Efficient hierarchical method for background subtraction. *Pattern Recognition*, pp. 2706-2715.

Yedidia, J. S., Freeman W. T., Weiss, Y., 2002. Understanding belief propagation and its generalize-tions. TR-2001-22.

Zhang, L., Seitz, S. M., 2007. Estimating optimal parameters for MRF stereo from a single image pair. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no.2, pp. 331-342.

McHugh, J. M., Konrad, J., Saligrama, V., Jodoin, P., 2009. Foreground-adaptive background subtraction. *IEEE Signal Processing Letters*, vol.16, issue 5, pp.390-393.

Xu, X., Huang, T. S., 2008. A loopy belief propagation approach for robust background estimation. *CVPR 2008*, pp. 23-28.