Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

GSM-MRF based classification approach for real-time moving object detection*

Xiang PAN, Yi-jun WU

(*Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China*)

E-mail: panxiang@zju.edu.cn; wuyijun1984@hotmail.com

Received May 23, 2007; revision accepted Aug. 21, 2007

Abstract: Statistical and contextual information are typically used to detect moving regions in image sequences for a fixed camera. In this paper, we propose a fast and stable linear discriminant approach based on Gaussian Single Model (GSM) and Markov Random Field (MRF). The performance of GSM is analyzed first, and then two main improvements corresponding to the drawbacks of GSM are proposed: the latest filtered data based update scheme of the background model and the linear classification judgment rule based on spatial-temporal feature specified by MRF. Experimental results show that the proposed method runs more rapidly and accurately when compared with other methods.

Key words: Moving object detection, Markov Random Field (MRF), Gaussian Single Model (GSM), Fisher Linear Discriminant Analysis (FLDA)

doi:10.1631/jzus.A071267 **Document code:** A **CLC number:** TP391.41

INTRODUCTION

Moving object detection with a fixed camera is a crucial issue in video processing, and is widely used for a variety of applications of computer vision, including object tracking and surveillance. Background subtraction based moving object detection is a method typically used to detect unusual motion in many situations, which heavily depends on the background model. A good background model should be robust to illumination changes, weather conditions, etc. Various background modeling algorithms have been developed, which include (1) pixel based operation like background appearance adaptation, threshold updating and pixel labelling, (2) region based operation which uses the statistical characteristics of a region, and (3) hierarchy-based model which processes the image at multi-resolution level.

In this paper, we propose a precise, stable and fast algorithm based on Gaussian Single Model (GSM), Markov Random Field (MRF) and Fisher Linear Discriminant Analysis (FLDA), modeling each pixel as a single Gaussian model and building an adaptive relationship network specified by MRF per pixel to judge background or foreground by FLDA. The rest of the paper is organized as follows. Section 2 reviews the related works in this field. The proposed approach is presented in Section 3. Experimental results are shown in Section 4, followed by our conclusions in Section 5.

PREVIOUS WORK

A popular method includes maintaining the background pixel intensity model, subtracting the background model from the new frame, and thresholding the difference value to judge background or foreground. However this approach always fails where the background is not completely stationary. For overcoming this drawback, many improved algorithms have been proposed, and their details are summarized below.

* Project (No. 10577017) supported by the National Natural Science Foundation of China

For pixel-based models, Pfinder (Wren *et al*., 1997) uses a single Gaussian background model (GSM) per pixel, it works very well in a slow scene change. A mixture of Gaussian models (GMM) (Stauffer and Grimson, 1999) is more powerful in practice when background is not completely statistic. However, GMM approaches require high computational complexity. Unlike parametric models, the non-parametric approaches estimate density functions directly from the sample data (Elgammal *et al*., 2002), which can adapt quickly to changes in the background.

The second category of methods use region models of the background. Eigen-space decomposition based object detection is proposed (Oliver *et al*., 2000; Seki *et al*., 2003), where the background is modeled by its principal components. Heikkila and Pietikainen (2006) used the Local Binary Pattern (LBP) texture operator to build a background model, while spatial MRF through Gibbs distribution has been widely used for background modeling (Paragios and Tziritas, 1999; Yaakov and Averbuch, 2001; Sheikh and Shah, 2005; Berrabah *et al*., 2006). This method is based on the construction of a global cost function to find the optimal label by ICM (Iterated Conditional Modes) and HCF (Highest Confidence First) or other algorithms.

Both spatial and temporal constraints are considered in multi-resolution based methods. Toyama *et al*.(1999) developed a three-component system for background maintenance. Meanwhile Zhou *et al*. (2005) proposed a time dependent pyramidal MRF to represent the state of the foreground and the background for each pixel in the pyramid.

PROPOSED METHOD

We have developed a novel approach to the processing of background subtraction based on the combination of GSM and MRF. After a typical GSM is analyzed, a new updating scheme is proposed to adapt to scene changes. Unlike previous MRF, we concentrate on the local constraint for each pixel, and construct a feature space based on an adaptive relationship network around its neighborhood system, which is specified by MRF, and then make the decision by FLDA. On one hand, it exploits the contextual information sufficiently when making decisions. On the other hand, it reduces computational costs to find the global optimal solution.

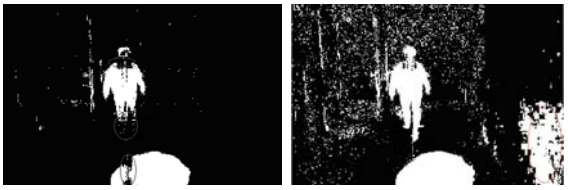
Analysis of general GSM

GSM method uses single Gaussian distribution to model the probability distribution of the each pixel intensity. Assuming the value of pixel at point (x,y) is $I(x,y)$, then $I(x,y)\sim N(u(x,y), \Sigma(x,y))$, where $u(x,y)$ and $\Sigma(x,y)$ are the mean and covariance, respectively. The probability of each pixel in the new image is calculated and compared with a pre-determined threshold h, and labelled as background (resp. foreground) if it is more (resp. less) than h. Then each parameter is updated by

$$\begin{cases} u(x,y;t) = (1-a)\cdot u(x,y;t-1) + a\cdot I(x,y;t), \\ \Sigma(x,y;t) = (1-a)\cdot \Sigma(x,y;t-1) + a\cdot d\cdot d^{\mathrm{T}}, \end{cases} \quad (1)$$

where $d=I(x,y;t)-u(x,y;t)$ denotes the distance between the new value and the mean value at point (x,y). a $(0\le a\le 1)$ is called update coefficient, which shows the update speed of the background model. The larger the a, the faster the update speed. a is always set to be '0' if the pixel is foreground and a small value for background pixels. In practice, it is difficult to choose a suitable value for a, for a small shift of a may lead to false decision. So could we find a way to update the background model without considering the update coefficient?

As for the threshold h, it usually does not yield a good result if a uniform threshold is chosen. Fig.1a and Fig.1b show the segmentation results with a high threshold and a low threshold, respectively. When h is too high, some foreground pixels are omitted, or some background pixels are detected as foreground ones. So it is unsuitable to use a uniform threshold.



(a) (b)

Fig.1 Results with different thresholds. (a) High threshold; (b) Low threshold

To overcome the drawbacks of the general GSM method, two main improvements are described in the following subsections.

Improved update scheme of GSM

The latest N background values are always used to calculate the mean and covariance to reflect the true distribution. So a regressive update scheme is proposed as

$$\begin{cases} u(x,y;t)=u(x,y;t-1)+[I(x,y;t)-I(x,y;t-N)]/N, \\ \Sigma(x,y;t)=-[I(x,y;t-N)-u(x,y;t-1)]^2/(N-1)+ \\ \qquad \Sigma(x,y;t-1)+[I(x,y;t)-u(x,y;t-1)]^2/(N-1), \end{cases} \tag{2}$$

where $u(x,y;t)$ and $\Sigma(x,y;t)$ stand for the mean and covariance at time t, respectively. In practice, some mistakes are unavoidable, so the data that will be used in the update scheme are first filtered to reduce the influence of mistakes and noises, and the filter is defined as

$$F=\begin{cases} 1, & |I(x,y;t)-u(x,y;t-1)|<2|\Sigma(x,y;t-1)|, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The above scheme remains sufficiently sensitive to the illumination change and is robust against noise.

MRF-based classification approach

To avoid choosing an arbitrary threshold and use contextual constraints, a classification based method in the feature space specified by MRF is proposed in this subsection.

Many problems in image analysis can be formulated as a labelling problem with contextual information, and MRF theory provides a convenient way of modeling context dependent entities such as image pixels and other spatially correlated features (Li, 1995). MRF assumes every image as a 2D lattice $S=\{s_1,s_2,\ldots,s_n\}$, and a set of possible labels for each site $f=\{f_1,f_2,\ldots,f_k\}$. For a given observation $D=\{d_1, d_2,\ldots,d_n\}$, the best label ($f^*$) corresponds to the maximum of posterior distribution of MRF: $p(f|d)$. According to the Bayes rule, the Hammersley-Chifford theorem and the equivalence between MRF and Gibbs distribution, the MAP-MRF (Geman and Geman, 1993) is

$$\begin{cases} f^*=\arg\max_f\{p(f\mid d)\}, \\ \propto \arg\min_f\{U(f\mid d)\}=\arg\min_f\{U(d\mid f)+U(f)\}, \end{cases} \tag{4}$$

where $U(f|d)$ is called posterior energy function, $U(f)$ and $U(d|f)$ are called prior and likely energy function, respectively. The above formula means that the best label corresponds to the minimum of posterior energy function. For moving object detection, it is a binary labelling problem, where $f=\{0,1\}$ and '1' stands for the foreground while '0' stands for the background.

The typical MRF-based moving object detection methods concentrate on the global posterior energy function (Paragios and Tziritas, 1999; Yaakov and Averbuch, 2001; Sheikh and Shah, 2005; Zhou *et al*., 2005). Since there are 2^{MN} (when image size is $M\times N$) possible configurations of f, it is not suitable for the applications for its heavy computational cost. Although many algorithms have been investigated, like ICM (Paragios and Tziritas, 1999), Gibbs Sample (Zhou *et al*., 2005), HCF (Yaakov and Averbuch, 2001), Graph cut (Sheikh and Shah, 2005), computational complexity and poor performance are still the main problems. In this paper, we convert the optimal label problem to a classification problem, which means that each pixel is classified independently in its spatial-temporal feature space specified by MRF.

First, we construct $U(f_i)$ and $U(d_i/f_i)$. Here, a spatial-temporal neighborhood system is defined, which is shown in Fig.2, including 8 neighborhoods in the current frame and 9 neighborhoods in the previous frame. The prior energy function $U(f_i)$ is defined as

$$U(f_i)=\sum\nolimits_{j\in C_i}\beta_{i,j}\cdot[1-2\delta(f_i-f_j)], \tag{5}$$

where C_i denotes the neighborhood system of point i, $\delta(f_i-f_j)$ is an impulse function and it is defined as

$$\delta(x)=\begin{cases} 1, & \text{if } x=0, \\ 0, & \text{otherwise.} \end{cases}$$

In Eq.(5), $\beta_{i,j}>0$ is a coefficient denoting the relativity between points i and j. The likelihood energy function is defined as

$$U(d_i \mid f_i)=\begin{cases}\dfrac{1}{2}\ln(2\pi\Sigma_i^2)+\dfrac{(d_i-u_i)^2}{2\Sigma_i^2}, & f_i=0,\\ -\ln Z, & f_i=1,\end{cases} \quad (6)$$

where the likelihood energy function equals the logarithm of the probability of a pixel in the background and each pixel has a uniform probability of being the foreground, such as Z=1/256 for gray images.

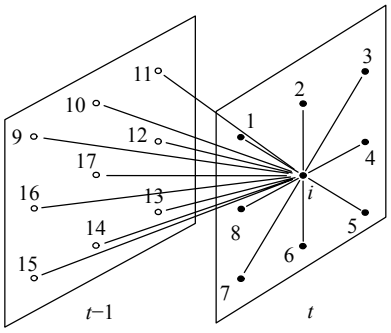


Fig.2 Spatial-temporal neighborhood system: 8 neighborhoods (1~8) in the current frame and 9 neighborhoods (9~17) in the previous frame

For each pixel, there are 17 parameters: $\beta_i=\{\beta_{i,1},\beta_{i,2},\ldots,\beta_{i,17}\}$. It is difficult to obtain all the values. We also notice that the neighborhood pixels with similar intensity have more consistency than others. Besides, the history label at the same position is more authentic than the other 16 neighborhood pixels. So for simplicity, two 8-neighborhoods on the current and previous frames and one neighborhood at the same position in the previous frame share the same coefficient, defined as $a_{i,1}$, $a_{i,2}$ and $a_{i,3}$, respectively, and every pixel of its neighborhood system is given a weight parameter $w_{i,j}$ to adjust the relativity among them, which is defined as

$$w_{i,j}=g\left((|\Delta I|+1)^{-1}\right),\ \ \beta_{i,j}=w_{i,j}\cdot a_{i,k(j)}, \quad (7)$$

where $k(j)$ depends on j, ΔI denotes the difference of intensity, $g(\Delta I)$ is inversely proportional to ΔI, and the $g(\Delta I)$ used in practice is defined as

$$g(\Delta I)=(1+\Delta I)^{-1/2}. \quad (8)$$

After modeling MRF, each pixel is classified by the following linear discriminant function, which is shown as follows:

$$\begin{aligned} f(a_{i,k(j)}) &= U(f_i \mid d_i)|_{f_i=1}-U(f_i \mid d_i)|_{f_i=0}\\ &= A_i a_{i,1}+B_i a_{i,2}+C_i a_{i,3}+D_i \\ &= \overline{\boldsymbol{a}}_i\cdot[A_i,B_i,C_i,D_i]^{\mathrm{T}}, \end{aligned} \quad (9)$$

where $\overline{\boldsymbol{a}}_i=[a_{i,1},a_{i,2},a_{i,3},1]$, and

$$\begin{cases} A_i=\sum_{j=1}^{8}2w_{i,j}a_{i,1}[\delta f_j-\delta(1-f_j)],\\ B_i=\sum_{j=9}^{16}2w_{i,j}a_{i,2}[\delta f_j-\delta(1-f_j)],\\ C_i=2w_{i,17}a_{i,3}[\delta f_{17}-\delta(1-f_{17})],\\ D_i=U(d_i \mid f_i)|_{f_i=1}-U(d_i \mid f_i)|_{f_i=0}. \end{cases} \quad (10)$$

Actually Eq.(9) stands for a two-class problem: the background class (f>0) and the foreground class (f<0).

Finally, the coefficient $\overline{\boldsymbol{a}}_i$ should be estimated. In Eq.(9), the vector $[A_i, B_i, C_i, D_i]^{\mathrm{T}}$ is a sample of a 4D space. Actually, the coefficient $[a_{i,1}, a_{i,2}, a_{i,3}, 1]$ describes a discriminant hyper plane for classifying $[A_i, B_i, C_i, D_i]^{\mathrm{T}}$ to be background class or foreground class. Assume that we have K frames to train the parameters. Fig.3 shows the distribution of the sample vector, which is linearly discriminative. So Linear Discriminant Analysis based method can be used here, and FLDA was used in the experiments.

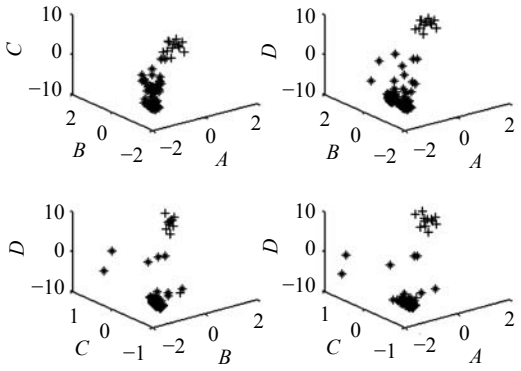


Fig.3 Distribution of the sample [A_i, B_i, C_i, D_i] showing [A_i, B_i, C_i], [A_i, B_i, D_i], [B_i, C_i, D_i], and [A_i, C_i, D_i], respectively. + and * stand for foreground and background, respectively

In the end, the proposed method can be summed up as two steps:

(1) Get an initialized label field by subtracting mean image of the GSM model from the current frame. Thus a threshold is needed to get the foreground candidates;

(2) The MRF based linear classification algorithm is implemented on the initialized result to delete the false candidates.

EXPERIMENTAL RESULTS

In order to validate the effectiveness of the proposed method, we tested some video sequences in gray images, including leaves strewing, moving curtain, illumination change and projection on the wall. The previous 30 and 100 frames were used to train the GSM model and discriminant function, respectively.

In the first step, a low threshold is used to ensure that most foreground pixels are detected as foreground candidates. Meanwhile, the threshold used in training and detection periods must share the same value. In our experiments, $2|\Sigma(x,y;t-1)|$ was used as the threshold.

The intuitionistic results obtained with the proposed method are shown in Fig.4. Some other methods were also tested simultaneously for comparison. For GSM method, the previous 30 frames are used to build the single Gaussian model, and a suitable threshold tested manually is used. For GMM method, the previous 30 frames are used to build 3 Gaussian models, whose parameters are estimated by *K*-means approximation, and the update coefficient of the background model and weight of each model are set to be 0.01 and 0.1, respectively. For PCA method, the image is divided into many blocks of 15×15, when the algorithm can run at the highest speed.

For the evaluation of the proposed method, we estimate the percentage of the pixels belonging to the foreground that are correctly labelled (percentage of true foreground detection, PTD) and the percentage of the background pixels that are incorrectly classified as foreground (percentage of false background detection, PFD), as shown in Fig.5, where the total number of detected foreground pixels are also presented, and the ground truth is labelled manually.

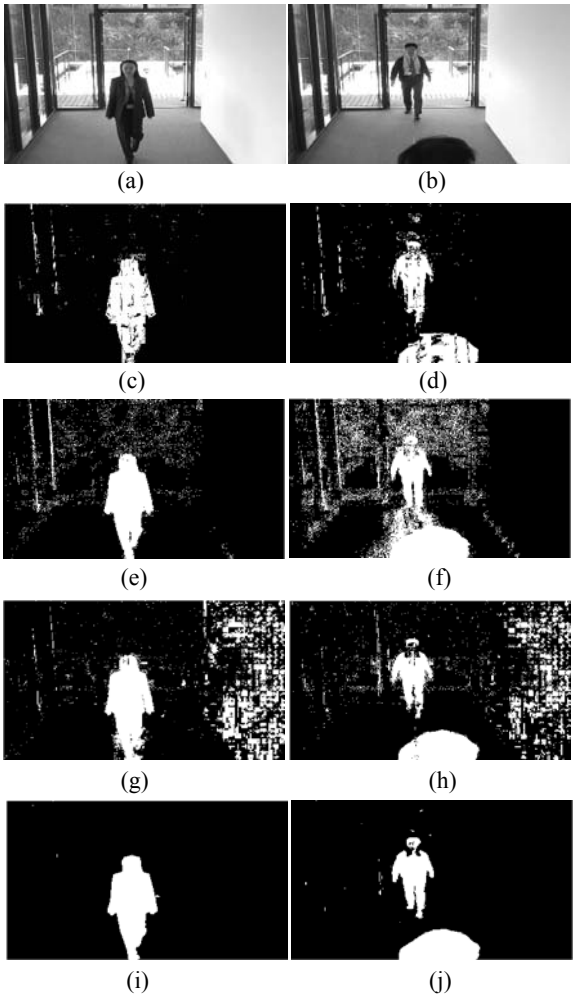


Fig.4 Detection results with different methods. (a) and (b) are the original images; (c), (e), (g) and (i) are the detection results of (a) by PCA, GMM, GSM and the proposed method, respectively; (d), (f), (h) and (j) are the detection results of (b) by PCA, GMM, GSM and the proposed method, respectively

When the parameters have been obtained, the proposed method can run at about 24 frames per second on 1.86 GHz Pentium processor and 512 M memory for 240×320 image, depending on the size of the detected foreground. This speed can satisfy almost all practical applications.

CONCLUSION

In this paper, we analyzed the performance of GSM and proposed two main improvements, including an update scheme for the background model and a linear judgment rule based on the spatial-

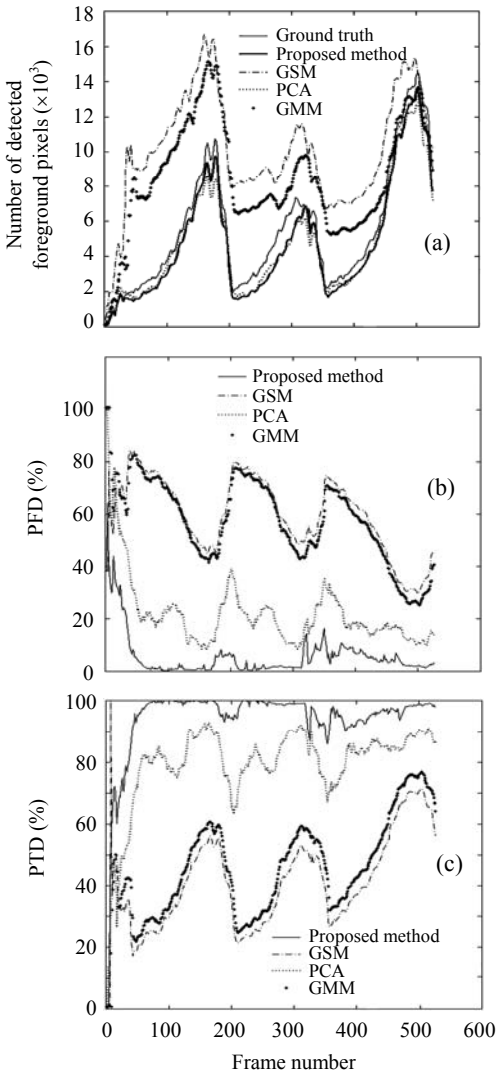


Fig.5 Performances of PCA, GMM, GSM and the proposed method. (a) Number of detected foreground pixels; (b) Percentage of false background detection (PFD); (c) Percentage of true foreground detection (PTD)

temporal feature specified by MRF, which overcomes the drawbacks of GSM. Unlike many other methods that use global optimal solution for MRF, we built a reliable relationship network for each pixel to get enough contexture information and reduce the processing time. Experiments show that the proposed method works better than GMM and PCA methods, and it runs fast enough to satisfy most online applications.

However, the currently proposed method works only on gray images, without considering the shadows. Adding the color information and characteristic of shadow to the feature space will be the main work in the future.

References

Berrabah, S.A., Cubber, G.D., Enescu, V., 2006. MRF-based Foreground Detection in Image Sequences from a Moving Camera. IEEE Int. Conf. on Image Processing, p.1125-1128.

Elgammal, A., Harwood, D., Davis, L., 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proc. IEEE*, **90**(7): 1151-1163. [doi:10.1109/JPROC.2002.801448]

Geman, S., Geman, D., 1993. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *J. Appl. Stat.*, **20**:25-62. [doi:10.1080/02664769300000058]

Heikkila, M., Pietikainen, M., 2006. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. on Pattern Anal. Machine Intell.*, **28**(4):657-662. [doi:10.1109/TPAMI.2006.68]

Li, S.Z., 1995. Markov Random Field Modeling in Computer Vision. Springer-Verlag.

Oliver, N.M., Rosario, B., Pentland, A.P., 2000. A Bayesian computer vision system for modeling human interactions. *IEEE Trans. on Pattern Anal. Machine Intell.*, **22**(8):831-843. [doi:10.1109/34.868684]

Paragios, N., Tziritas, G., 1999. Adaptive detection and localization of moving objects in image sequences. *Signal Processing: Image Commun.*, **14**:277-296. [doi:10.1016/S0923-5965(98)00011-3]

Seki, M., Wada, T., Fujiwara, H., Sumi, K., 2003. Background Detection Based on the Cooccurrence of Image Variations. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, **2**:65-72.

Sheikh, Y., Shah, M., 2005. Bayesian modeling of dynamic scenes for object detection. *IEEE Trans. on Pattern Anal. Machine Intell.*, **27**(11):1778-1792. [doi:10.1109/TPAMI.2005.213]

Stauffer, C., Grimson, W.E.L., 1999. Adaptive Background Mixture Models for Real-Time Tracking. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, **2**:246-252.

Toyama, K., Krumm, J., Brumitt, B., Meyers, B., 1999. Wallower: Principles and Practice of Background Maintenance. IEEE Conf. on Computer Vision, p.255-261.

Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P., 1997. Pfinder: real-time tracking of the human body. *IEEE Trans. on Pattern Anal. Machine Intell.*, **19**(7):780-785. [doi:10.1109/34.598236]

Yaakov, T., Averbuch, A., 2001. A Region-based MRF Model for Unsupervised Segmentation of Moving Objects in Image Sequences. Computer Society Conf. on Computer Vision and Pattern Recognition, **1**:889.

Zhou, Y., Xu, W., Tao, H., Gong, Y.H., 2005. Background Segmentation Using Spatial-Temporal Multi-Resolution MRF. IEEE Workshop on Motion and Video Computing, **2**:8-13. [doi:10.1109/ACVMOT.2005.32]