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

Bayesian moving object detection in dynamic scenes using an adaptive foreground model*

Sheng-yang YU†, Fang-lin WANG, Yun-feng XUE, Jie YANG

(*Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China*)

†E-mail: whizstorm@sjtu.edu.cn

Received Oct. 24, 2008; Revision accepted Feb. 10, 2009; Crosschecked Aug. 14, 2009; Published online Oct. 21, 2009

Abstract: Accurate detection of moving objects is an important step in stable tracking or recognition. By using a nonparametric density estimation method over a joint domain-range representation of image pixels, the correlation between neighboring pixels can be used to achieve high levels of detection accuracy in the presence of dynamic background. However, color similarity between foreground and background will cause many foreground pixels to be misclassified. In this paper, an adaptive foreground model is exploited to detect moving objects in dynamic scenes. The foreground model provides an effective description of foreground by adaptively combining the temporal persistence and spatial coherence of moving objects. Building on the advantages of MAP-MRF (the maximum a posteriori in the Markov random field) decision framework, the proposed method performs well in addressing the challenging problem of missed detection caused by similarity in color between foreground and background pixels. Experimental results on real dynamic scenes show that the proposed method is robust and efficient.

Key words: Moving object detection, Foreground model, Kernel density estimation (KDE), MAP-MRF estimation
doi:10.1631/jzus.A0820743 **Document code:** A **CLC number:** TP391.41

INTRODUCTION

In typical automated surveillance applications, a stationary camera is usually used to monitor a scene. Accurate detection of moving objects is an important task in these applications, where background subtraction is widely adopted. However, the setup of a stationary camera does not necessarily guarantee a stationary background. Examples of non-stationary backgrounds abound in the real world, including periodic motions, such as a ceiling fans or escalators, and dynamic textures, such as waving trees or flowing water. Furthermore, the camera will not keep absolutely stationary due to wind or vibration in the environment, which can also lead to background motion. Various statistical background modeling techniques have been introduced to model the uncertainties of background pixels for moving object detection (Wren *et al.*, 1997; Stauffer and Grimson, 2000; Elgammal *et al.*, 2002).

The idea that the color of a pixel over time in a static scene could be modeled by a single Gaussian distribution was introduced by Wren *et al.*(1997), who modeled the color of each pixel with a single 3D Gaussian. Stauffer and Grimson (2000) proposed the Gaussian mixture model (GMM) to model the multimodality of the underlying background probability density function. The decision whether a pixel belongs to the background is made by comparing it with each Gaussian density. Nonparametric data-driven kernel density estimation (KDE) was used by Elgammal *et al.*(2002) to address the uncertainty of spatial location and to handle multiple modes in the intensity of the background.

These algorithms model each pixel independently. They ignore the significant correlation that exists between the intensities of neighboring pixels,

* Project (Nos. 60602012 and 60675023) supported by the National Natural Science Foundation of China, the National High-Tech Research and Development Program (863) of China (No. 2007AA01Z 164), and the Shanghai Key Laboratory Opening Plan Grant (No. 06dz22103), China

which can be used to achieve higher detection accuracy in the presence of dynamic background. To address the issue, Sheikh and Shah (2005) modeled background as a single probability density distribution using a nonparametric KDE method over a joint domain-range representation of image pixels. Furthermore, they explicitly modeled the foreground in a fashion consistent with the background model to augment the detection of objects. Finally, instead of directly applying a threshold to membership probabilities, a MAP-MRF (the maximum a posteriori in the Markov random field) decision framework was proposed in which the background and foreground models were combined in a Bayesian framework. It has been shown that joint domain-range scene modeling based on nonparametric KDE is better than the previously proposed methods. Sheikh's moving object detection scheme performs suitably well in several challenging situations.

However, in real world scenes the observed color features generated by the foreground and background often overlap strongly in the same position (Fig.1). Most existing statistical object detection algorithms (Wang and Tan, 2002; Mahamud, 2006; Sun *et al.*, 2006; Lu and Hager, 2007) present a weak description of foreground. Though Sheikh and Shah (2005) used the temporal persistence of moving objects to model the foreground, many foreground pixels are still misclassified when their color distribution is similar to that of the background pixels. A large number of misclassified pixels, which cannot be simply repaired by morphological operators, will destroy the structure of foreground objects and inevitably affect the consequent object recognition or analysis.

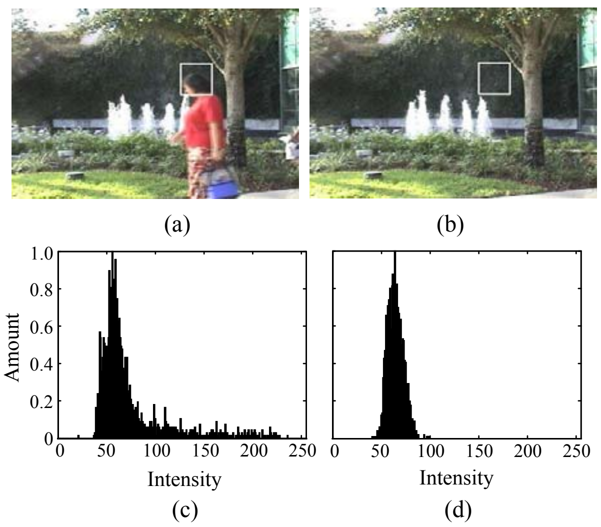

Fig.1 An example of foreground and background color overlapping in a dynamic scene
(a) A frame of the fountain sequence; (b) A background image of (a); (c) and (d) are intensity histograms of the region within white rectangles (41×41) in (a) and (b), respectively

Since the missed foreground detection is far more severe than false foreground detection using the energy minimization framework (Sheikh and Shah, 2005), the motivation for our work is to improve the detection of foreground in situations where the foreground and background have similar color distributions and in the meantime prevent any increase in false foreground detection. In this paper, an accurate foreground modeling method is proposed, which adaptively combines the temporal persistence and spatial coherence of moving objects. Then, the adaptive foreground model is competitively used in a MAP-MRF decision framework to detect moving objects in dynamic scenes. Definitive experiments show that the proposed algorithm can model foreground adaptively and achieve accurate detection of moving objects in dynamic scenes.

BAYESIAN MOVING OBJECT DETECTION FORMULATION

For a video sequence taken by a fixed camera, each frame image contains N pixels. Let S be the set of indices referring to each of the N pixels. Given a set of observations $I=\{\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_i, \ldots, \mathbf{I}_N\}$, $i\in S$ of the current frame at time-step t, the task of object detection is to assign a label $l_i\in\{0$ (background), 1 (foreground)$\}$ to each pixel $i\in S$, and obtain $l=\{l_1, l_2, \ldots, l_i, \ldots, l_N\}$.

In most published studies, object detection was attempted by first modeling the conditional distribution $p(\mathbf{I}_i|l_i)$ of feature value $\mathbf{I}_i$ at each pixel i independently. The model used can be either parametric (Wren *et al.*, 1997; Stauffer and Grimson, 2000) or nonparametric (Elgammal *et al.*, 2002; Sheikh and Shah, 2005) based on a past window of observed feature values at the given pixel. The background and foreground models will be detailed presently. Assume that the observed feature value of image pixels is conditionally independent given l, thus,

$$p(I\,|\,l)=\prod_{i=1}^{N}p(\boldsymbol{I}_i\,|\,l_i). \tag{1}$$

However, it is clear that neighboring labels are strongly dependent on each other. The neighborhood consistency can be modeled with a Markov random field (MRF) prior on the labels:

$$P(l)\propto\prod_{i=1}^{N}\prod_{j\in\varepsilon_i}\varphi(i,j), \tag{2}$$

$$\varphi(i,j)=\exp\left(\lambda(l_il_j+(1-l_i)(1-l_j))\right), \tag{3}$$

where λ determines the pair-wise interaction strength among neighbors and ε_i is the four-neighborhood of pixel i.

Given the MRF prior and the likelihood model above, moving object detection in a given frame reduces to a maximum a posteriori (MAP) $P(l|I)$ solution. According to the Bayes rule, the posterior is equivalent to

$$P(l\,|\,I)=\frac{\prod_{i=1}^{N}p(\boldsymbol{I}_i\,|\,l_i)\exp\left(\sum_{i=1}^{N}\sum_{j\in\varepsilon_i}\lambda\left(l_il_j+(1-l_i)(1-l_j)\right)\right)}{P(I)}, \tag{4}$$

where $P(I)$ is the density of I, which is a constant when I is given. Finally, the MAP estimate is the binary image that maximizes function Eq.(4). The discrete cost function Eq.(4) can be solved for the global optimum using standard graph-cut algorithms (Greig *et al.*, 1989; Kolmogorov and Zabih, 2004).

OBJECT DETECTION METHOD USING AN ADAPTIVE FOREGROUND MODEL

In the following subsections, the foreground and background models in Sheikh and Shah (2005) are reviewed and then the proposed foreground model is presented. Finally the complete object detection schedule is introduced.

Joint domain-range based KDE model

In most previous work, only the background feature distribution $p(\boldsymbol{I}_i|l_i=0)$ has been explicitly modeled using a past window of observed feature values that were labeled background in previous frames. To model the foreground, a uniform threshold value is often implicitly assumed, i.e., $p(\boldsymbol{I}_i|l_i=1)=\gamma$.

$$\gamma=\begin{cases}1/(NC), & \text{joint domain-range based KDE model,}\\ 1/C, & \text{range based KDE model,}\end{cases} \tag{5}$$

where N is the number of pixels of a frame and C is the number of possible color values. Sheikh and Shah (2005) defined γ as $1/(NC)$, which means that at any time a foreground pixel can have any color at any location in the image with a uniform probability. However, this foreground model does not take into account the knowledge gained from the objects detected in previous frames. Their foreground is also modeled in exactly the same manner as the background using a past window of observations that were labeled foreground in the previous frames.

The observation of image pixels is represented by $\boldsymbol{I}_i\in\mathbb{R}^5$, $i\in S$. The feature vector $\boldsymbol{I}_i$ is a joint domain-range representation, where the space of the image lattice, (x, y), is domain, and some color space, for instance (r, g, b), is the range. By adopting the joint domain-range representation, a single KDE model can be used for the entire background or foreground $f_{R,G,B,X,Y}(r, g, b, x, y)$.

The background model is built from the sample set $\psi_\mathrm{b}=\{\boldsymbol{I}_1^\mathrm{b},\boldsymbol{I}_2^\mathrm{b},...,\boldsymbol{I}_n^\mathrm{b}\}$, which contains all the background pixels. The parameter K_b ensures that the algorithm remains adaptive to slower change and corresponds to the learning rate of the detection algorithm. A kernel density estimator is built by assigning an appropriate kernel to each of these samples (Parzen, 1962). The probability that estimation point $\boldsymbol{I}_i$ belongs to the background is given as

$$p(\boldsymbol{I}_i\,|\,l_i=0)=p(\boldsymbol{I}_i\,|\,\psi_\mathrm{b})=\frac{1}{n}\sum_{j=1}^{n}\varphi_{\boldsymbol{H}}(\boldsymbol{I}_i-\boldsymbol{I}_j^\mathrm{b}). \tag{6}$$

Here, $\varphi_{\boldsymbol{H}}(\boldsymbol{X})=|\boldsymbol{H}|^{-1/2}\varphi(\boldsymbol{H}^{-1/2}\boldsymbol{X})$, φ is a d-variate kernel function, and $\boldsymbol{H}$ is a symmetric positive definite $d\times d$ bandwidth matrix (Wand and Jones, 1995). The d-variate Gaussian density with zero mean and unity variance is used as the kernel φ. To reduce computation, it is assumed that matrix $\boldsymbol{H}$ is diagonal, $\boldsymbol{H}=\mathrm{diag}\left\{\sigma_\mathrm{R}^2,\sigma_\mathrm{R}^2,\sigma_\mathrm{R}^2,\sigma_\mathrm{D}^2,\sigma_\mathrm{D}^2\right\}$.

The foreground model is constructed in a fashion consistent with the background model: a domain-range nonparametric density $\psi_{\mathrm{f}}=\{\boldsymbol{I}_1^{\mathrm{f}}, \boldsymbol{I}_2^{\mathrm{f}}, ..., \boldsymbol{I}_m^{\mathrm{f}}\}$, contains all the object pixels that appeared in the last K_{f} frames before time-step t.

$$p(\boldsymbol{I}_i \mid \psi_{\mathrm{f}}) = \frac{1}{m}\sum_{j=1}^{m} \varphi_{\boldsymbol{H}}(\boldsymbol{I}_i - \boldsymbol{I}_j^{\mathrm{f}}). \tag{7}$$

Once a foreground region is detected in the previous frame, there is an increased probability of observing a foreground region in the current frame, in the same proximity and with a similar color distribution. However, the uniform model can be used to detect objects which newly appear in the scenes. Thus, foreground probability is expressed as a mixture of the uniform value and the kernel density function:

$$p(\boldsymbol{I}_i \mid l_i = 1) = \alpha\gamma + (1-\alpha)p(\boldsymbol{I}_i \mid \psi_{\mathrm{f}}), \tag{8}$$

where α<<1 is the mixture weight with a fixed value.

Novel foreground model

Sheikh and Shah (2005) explicitly modeled the foreground with the mixture of a uniform distribution and the kernel density function to augment the detection of objects. When the object pixels and the background pixels have a similar color, the uniform foreground model (Eq.(5)) or the mixture foreground model (Eq.(8)) cannot model the foreground discriminatively. When the foreground model and background model are used competitively, many false detections are usually produced. Figs.2a and 2b show the detection result of Sheikh's algorithm and the uniform foreground model for Fig.1a. It is clear that the head and left hand of the walker are seriously misclassified, because of their similar color distribution.

To solve this problem, we introduce the spatial color feature of foreground objects in the current frame to the foreground model. The theory that neighboring pixels belonging to foreground should have similar color in a frame is used in our approach. Thus, if the color statistics of the nearby foreground pixels can be estimated, the constraint of spatial color can be used in the foreground model to enhance its discrimination.

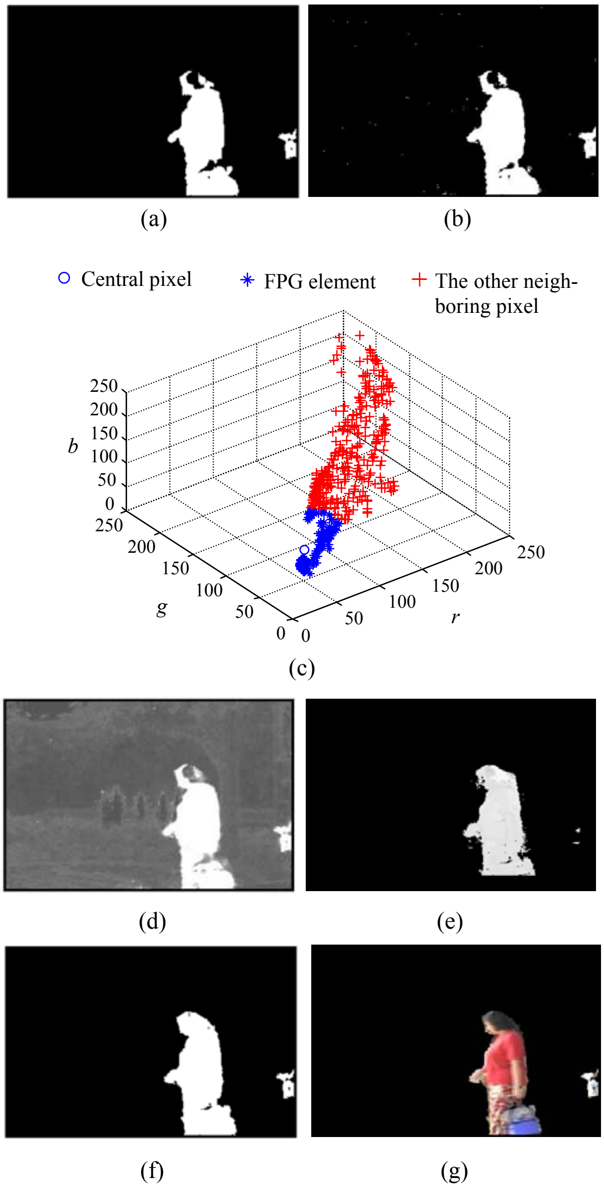

Fig.2 (a) Detection result using Sheikh's algorithm; (b) Result obtained using the uniform foreground model; (c) Determination of FPG (foreground peer group) for the pixel at the center of Fig.1a; (d) Background negative log-likelihood map; (e) Foreground likelihood map of the FPG based Gaussian foreground model; (f) Result obtained by the proposed method; (g) Original masked image

The spatial color model of a given pixel is learned from the nearby foreground pixels, which are identified by a simple likelihood ratio classifier. With the KDE based joint domain-range background model, we can mark the pixel that has a low log-likelihood ratio value as 'definite foreground'. The classifier is

$$\delta(i)=\begin{cases}1, & -\ln\dfrac{p(\boldsymbol{I}_i\mid\psi_{\mathrm{b}})}{\gamma}>\kappa,\\ 0, & \text{otherwise},\end{cases} \tag{9}$$

where κ is the threshold and is set to zero, which means the background is less likely than the foreground. Fig.2b is the result of this classifier. Let F denote the set of pixels marked as 'definite foreground' in the current frame, $F=\{i|i\in S, \delta(i)=1\}$. Let N_i be the set of pixels in a $w\times w$ window centered at pixel i. Therefore, $F_i=F\cap N_i$ is defined as the neighboring foreground pixels of pixel i. Then, the foreground peer group (FPG) F_i^{PG} can be constructed which consists of pixels with similar colors to the pixel i in F_i. The concept of a peer group was proposed by Deng *et al.*(1999). They used the peer group pixels to estimate the filtering result and a Fisher's discriminant estimation based approach to construct the peer group for each pixel. For computational reduction, the construction of a peer group can be simplified to finding pixels whose color values are close enough to the central pixel. Namely, $F_i^{\mathrm{PG}}=\{j\mid j\in F_i,\ \|\tilde{\boldsymbol{I}}_i-\tilde{\boldsymbol{I}}_j\|<\tau\}$. Here, $\tilde{\boldsymbol{I}}_i$ is a 3D color vector of pixel i, ignoring the domain information of feature vector $\boldsymbol{I}_i$. Fig.2c shows the determination of FPG for the central pixel of the rectangle region in Fig.1a in RGB color space. τ is set to 60.

To calculate the foreground probability of pixel i in the current frame using the local foreground color distribution, a Gaussian distribution is used to model the local foreground color over the FPG, i.e., $p_{\mathrm{FPG}}(\tilde{\boldsymbol{I}}_i)=N(\tilde{\boldsymbol{I}}_i\mid\mu_i,\Sigma_i)$.

The FPG-based Gaussian foreground model can handle the local color coherence of objects effectively, but it will also bring some false foreground detection. However, the main problem of the joint domain-range based KDE foreground model is the low discriminative capability of similar color distributions. Therefore, our foreground model approach is to combine the temporal persistence and the spatial coherence of moving objects adaptively:

$$p(\boldsymbol{I}_i\mid l_i=1)=\alpha_i' p_{\mathrm{FPG}}(\tilde{\boldsymbol{I}}_i)+(1-\alpha_i')p(\boldsymbol{I}_i\mid\psi_{\mathrm{f}}), \tag{10}$$

where α_i' is the mixture weight, which is a trade-off between the FPG-based Gaussian model and the joint domain-range based KDE model. If the foreground and background colors can be well separated, it should rely more on the second term; otherwise, it should rely on the first term. For this purpose, we adaptively mix two models based on the discriminative capabilities of the joint domain-range based foreground and background models. Our adaptive mixture is

$$\alpha_i'=\exp\left(-\left|\ln\frac{p(\boldsymbol{I}_i\mid\psi_{\mathrm{f}})}{p(\boldsymbol{I}_i\mid\psi_{\mathrm{b}})}\right|\right). \tag{11}$$

If the foreground and background colors can be well separated, i.e., the likelihood difference is large, the mixture weight α_i' is set to be small, and the foreground model is set to rely more on the joint domain-range based KDE model. Otherwise, α_i' is large and the foreground model relies more on the FPG-based Gaussian model.

Fig.2d shows the background negative log-likelihood map of Fig.1a. Fig.2e is the foreground log-likelihood map of the FPG-based Gaussian foreground model.

Object detection schedule

Instead of using only likelihoods, prior information of neighborhood spatial context is enforced in a MAP-MRF framework. Sheikh and Shah (2005) provided a good object detection scheme which we have also used. The complete object detection algorithm consists of a detection step and a model update step, described as follows:

Detection step: compute $p(\boldsymbol{I}_i|l_i=0)$ (Eq.(6)) and $p(\boldsymbol{I}_i|l_i=1)$ (Eq.(10)), and then maximize Eq.(4) by the graph-cut algorithm.

Model update step: remove all pixels in ψ_{f} from K_{f} frames ago and append all pixels detected as foreground in the current frame to ψ_{f}, remove all pixels in ψ_{b} from K_{b} frames ago, and append all pixels of the current frame to ψ_{b}.

Fig.2f is the result obtained by the proposed method and Fig.2g is the original masked image. The problem of misclassified regions of the walker is resolved, and no obvious false foreground detection is produced.

EXPERIMENTAL RESULTS

The algorithm was tested on three sequences with dynamic scenes on a PC with 2.8 GHz CPU and 1 GB RAM. Our C programming language implementation of the proposed approach using the adaptive foreground model can process about 3 to 4 frames/s for a frame size of 360×240.

For all the results, the bandwidth matrix $\boldsymbol{H}$ was parameterized as a diagonal matrix with three equal variances pertaining to the range, represented by σ_{R}^2 and two equal variances pertaining to the domain, represented by σ_{D}^2. The value used in the experiment was $(\sigma_{\mathrm{R}}^2, \sigma_{\mathrm{D}}^2) = (16, 25)$. The parameters K_{b} and K_{f} were set to 20 and 3, respectively. The size of the neighborhood window could be evaluated based on the bounding rectangle obtained by the coarse detected result with Eq.(9), and w was set to 41 for the first sequence and 21, 31, and 41 for the second and third sequences. Because of the adaptive mixing factor of Eq.(10), our foreground model can adjust the rate between the FPG-based Gaussian model and joint domain-range based KDE model. If the neighborhood of the calculating FPG is large enough to gain the spatial coherence constraint, the proposed foreground model is not sensitive to w.

The threshold τ which defines the maximum distance of color between pixels of FPG was set to 60 in the experiments. It is stressed that no morphological operators were used in the experiment. The experimental results for the proposed method, the GMM, and Sheikh's method are compared qualitatively and quantitatively.

Qualitative analysis

The fountain sequence (Sheikh and Shah, 2005) contains three sources of dynamic textures: tree branch oscillation, fountains, and the shadow of the tree on the grass below. Three frames typical of the sequence are shown in the top row of Fig.3. The experimental results for the proposed method, the GMM, and Sheikh's method are also shown in Fig.3.

The results of the GMM (Fig.3, second row) contain not only missed foreground detection but also false foreground detection. With this method, it is not easy to keep a balance between missed and false detection in a dynamic scene like the fountain sequence. The results of Sheikh's algorithm are consistent with previous analyses. Many misclassified foreground pixels are generated because they have a color distribution similar to the background. The misclassified foreground pixels (Fig.3, fourth row) are corrected while preventing an increase in the false foreground detection.

Fig.3 The top row shows the original images of the fountain sequence, the second row shows the results obtained using a five-component Gaussian mixture model, and the third row shows the results obtained using Sheikh's algorithm. The fourth row shows the results obtained by the proposed method, and the fifth row shows the original masked image

Fig.4 shows three successive frames and corresponding results of the car sequence, where a walker is moving in front of a similar-colored car. The object region is clearly split using Sheikh's algorithm. The proposed algorithm is able to correct most missed foreground detection and obtain an integrated object. The results also show that our method is not sensitive to the parameter w.

Fig.5 shows three successive frames and corresponding results of the hall sequence, where the challenge is the similar white color between the garment and the wall. The moving object can also be detected accurately by the proposed method.

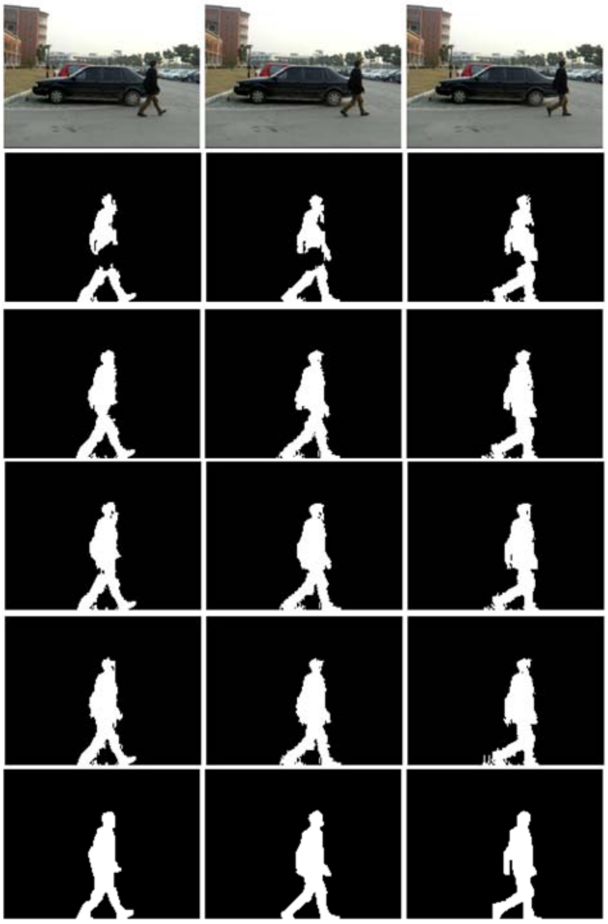

Fig.4 The top row shows three successive images of the car sequence, the second row shows the results obtained using Sheikh's algorithm, and the third to fifth rows show the results obtained by the proposed method with *w*=21, 31, and 41, respectively. The last row shows the ground truth images

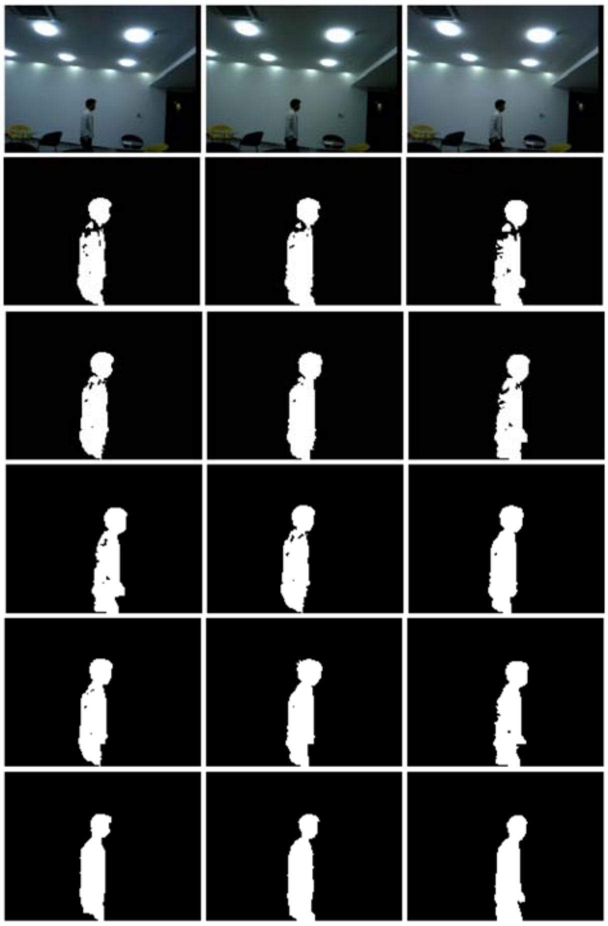

Fig.5 The top row shows three successive images of the hall sequence, the second row shows the results obtained using Sheikh's algorithm, and the third to fifth rows show the results obtained by the proposed method with *w*=21, 31, and 41, respectively. The last row shows the ground truth images

Quantitative analysis

To quantitatively evaluate the detection performance, we introduce precision and recall as

$$\text{Precision}=\frac{\text{Number of true positives detected}}{\text{Total number of positives detected}},$$

$$\text{Recall}=\frac{\text{Number of true positives detected}}{\text{Total number of true positives in ground truth}}.$$

We manually segmented 30 successive frames of the car and hall sequences into foreground and background as ground truth data. The per-frame detection rates of the proposed algorithm and Sheikh's algorithm are shown in Figs.6 and 7. The detection recall and precision at each level of the proposed approach are also shown. The detection using only the background model is equal to Eq.(9). The log-likelihood ratio using both the background model and the proposed foreground model is

$$L(i)=-\ln\frac{p(\boldsymbol{I}_i\,|\,l_i=1)}{p(\boldsymbol{I}_i\,|\,l_i=0)}.$$

If $L(i)>0$, i is labeled as foreground; otherwise, it is labeled as background.

With the car and hall sequences, the proposed method achieves detection accuracy similar to Sheikh's algorithm in terms of precision, and achieves higher detection accuracy in terms of recall. In other words, the missed foreground detection rate drops without any increase in false foreground detection.

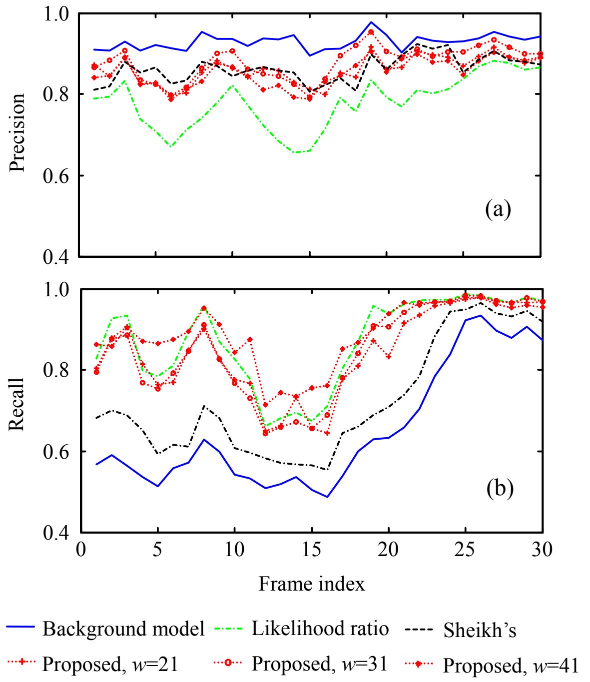

Fig.6 Detection precision (a) and recall (b) for the car sequence

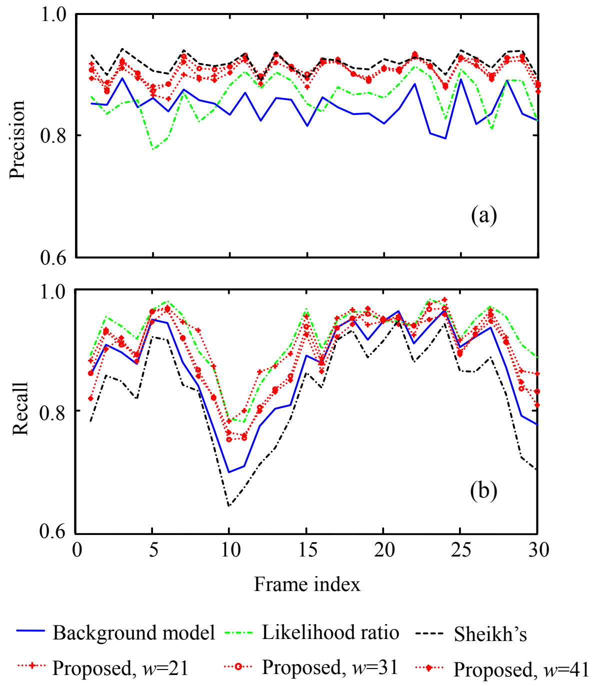

Fig.7 Detection precision (a) and recall (b) for the hall sequence

DISCUSSION AND CONCLUSION

In this paper, a foreground model is proposed that adaptively combines the temporal persistence and spatial coherence of a moving object. Building on the advantages of Sheikh's moving object detection framework, the proposed method performs well in the challenging situation of similar color in foreground and background pixels. Definitive experiments show that the proposed algorithm can model foreground effectively and improve the detection of objects while preventing any increase in false detection.

ACKNOWLEDGEMENTS

The authors are very grateful to Xiang Zhang for data support and the anonymous reviewers for their valuable suggestions.

References

Deng, Y., Kenney, C., Moore, M.S., Manjunath, B.S., 1999. Peer Group Filtering and Perceptual Color Image Quantization. Proc. IEEE Int. Symp. on Circuits and Systems, **4**:21-24. [doi:10.1109/ISCAS.1999.779933]

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]

Greig, D., Porteous, B., Seheult, A., 1989. Extract maximum a posteriori estimation for binary images. *J. Royal Stat. Soc. Ser. B*, **51**(2):271-279.

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**(2):147-159. [doi:10.1109/TPAMI.2004.1262177]

Lu, L., Hager, G.D., 2007. A Nonparametric Treatment for Location/Segmentation Based Visual Tracking. IEEE Conf. on Computer Vision and Pattern Recognition, p.1-8. [doi:10.1109/CVPR.2007.382976]

Mahamud, S., 2006. Comparing Belief Propagation and Graph Cuts for Novelty Detection. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, p.1154-1159. [doi:10.1109/CVPR.2006.80]

Parzen, E., 1962. On estimation of a probability density and mode. *Ann. Math. Statist.*, **33**(3):1065-1076. [doi:10.1214/aoms/1177704472]

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

Stauffer, C., Grimson, W., 2000. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach.*

Intell., **22**(8):747-757. [doi:10.1109/34.868677]
Sun, J., Zhang, W., Tang, X., Shum, H.Y., 2006. Background cut. *LNCS*, **3952**:628-641. [doi:10.1007/11744047_48]
Wand, M., Jones, M., 1995. Kernel Smoothing. *In*: Cox, D.R., Hinkley, D.V., Reid, N., *et al*. (Eds.), Monographs on Statistics and Applied Probability. Chapman and Hall, New York.
Wang, Y., Tan, T., 2002. Adaptive Foreground and Shadow Detection in Image Sequences. Proc. Int. Conf. on Pattern Recognition, p.983-986.
Wren, C., Azarbayejani, A., Darrel, T., Pentland, A.P., 1997. Pfinder: real time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**(7):780-785. [doi:10.1109/34.598236]