Journal of Southeast University (English Edition) Vol. 20 No. 4 Dec. 2004 ISSN 1003—7985

Multimodal background model with noise and shadow suppression for moving object detection

Mao Yanfen Shi Pengfei

(School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: A statistical multimodal background model is described for moving object detection in video surveillance. The solution to some of the problems such as illumination changes, initialization of model with moving objects, and shadows suppression is provided. The background samples are chosen by thresholding inter-frame differences, and the Gaussian kernel density estimation is used to estimate the probability density function of background intensity. Pixel's neighbor information is considered to remove noise due to camera jitter and small motion in the scene. The hue-max-min-diff color information is used to detect and suppress moving cast shadows. The effectiveness of the proposed method in the foreground segmentation is demonstrated in the traffic surveillance application.

Key words: video surveillance; background model; kernel density estimation; shadow suppression; hue-max-min-diff (HMMD) color space

Video surveillance systems aim to automatically detect objects in different kinds of environments. Background subtraction is a common module for differentiating background from foreground, and those foreground pixels should be processed for high-level analysis such as classification or tracking. The difficulty of background subtraction technique is mainly in the maintenance and update of the scene representation, which is called the background model.

In the literature, many works have been published on background modeling. Tab.1 provides a comparison of some popular methods ranging from model modality. The unimodal approaches[1-4] assume the probability density function of the pixel feature can be modeled with a single modal distribution. They can adapt to slow changes in the scene, e.g., gradual illumination changes, by recursively updating the model using a simple adaptive filter. However, in an outdoor cluttered environment, the scene background is not completely static; a single modal assumption will not hold. Some approaches are proposed to model the background with multimodal models[5-8]. Stauffer and Grimson extended the unimodal model (the threshold T is smaller) to multimodal (T is larger) by modeling the pixel color as a mixture of Gaussians (MoG). The multimodal models can handle a background that is not completely static and keep several modals in the background, so the instantaneous background cannot be represented by a single image.

Tab.1 Comparison of modeling approaches

Modality	Methods
Unimodal	Mean
	Median
	Single Gaussian[1,2]
	Pfinder[3]
	Friedman[4]
	Mixture of Gaussian[5] (T is smaller)
Multimodal	W4 (Bimodal)[6]
	Mixture of Gaussian[5] (T is larger)
	Wallflower[7]
	Elgammal[8]
	Our approach

Background maintenance in itself is application oriented. In the traffic surveillance application, some problems will be encountered as follows.

1) Bootstrapping Many methods assume an initial model obtained by using a training sequence in which no foreground objects are present. This puts serious limitations on systems to be used in high traffic areas, and it is necessary to train the model using a sequence containing moving objects.

2) Illumination change Scene illumination changes gradually at different times of day and suddenly due to a light switch. The background model should adapt to the variation to decrease the false positive.

3) Background motion Global motion is caused by small camera displacements. Local motion of tree branch sways with the wind. The model should be robust with respect to these motions.

Received 2004-05-09.

Foundation item: The National Basic Research Program of China (973 Program) (No. TG1998030408).

Biographies: Mao Yanfen (1975—), female, graduate; Shi Pengfei (corresponding author), male, professor, pfshi@sjtu.edu.cn.

4) Shadows Moving objects cast shadows that should not be classified as foreground. Static shadows are included in the background and will not cause serious problem.

To solve the above problems, a statistical model with kernel density estimation (KDE) is proposed. The thresholding of inter-frame difference is utilized to choose the background samples. The perceptual uniform and computation inexpensive hue-max-min-diff (HMMD) color space[9] is used to suppress the shadows.

1 Statistical Background Model

Several assumptions are given: ① The camera is stationary but with small jitter; ② The scene is approximately static and only small motion may occur; ③ The scene illumination may change; ④ Each pixel in the image will reveal the background for at least a short interval of the training sequence.

1.1 KDE and background subtraction

The values of a particular pixel over time can be considered as a pixel process, i.e. a time series of scalars for gray-value or vectors for color pixel values. Therefore, the background intensity/density probability can be estimated by KDE. The background sample set is extracted from the training sequence including the foreground objects. We use the thresholding results of inter-frame difference as the coarse background samples. This differs from the method of Elgammal, et al. [8], which chooses the samples directly from the training data and inevitably consists of the foreground points and results in many false negatives.

If the inter-frame difference is less than a certain threshold T_g, the newer intensity is chosen as the sample. Given N frames, the pixel in which inter-frame difference is larger than T_g is not exploited in the density estimation.

$$\begin{aligned} &s_i(x, y) = g_{i+n}(x, y), \\ &|g_i(x, y) - g_{i+n}(x, y)| < T_g \\ &\quad i = 1, 2, \cdots, N-n \end{aligned} \tag{1}$$

where n is the interval of frame, $g_i(x, y)$ is the intensity of pixel (x, y) at frame i, $s_i(x, y)$ is the background sample.

Let $S = \{s_1, \cdots, s_K\}$ be a sample set of pixel(x, y). $K \equiv \bar{K}(x, y)$ is the sample number. The density of background points can be estimated by the Gaussian kernel density function to model the multimodal background. The probability of the current pixel at time t belonging to background can be computed by

$$p_t \equiv p(g_t(x, y)) = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{\sqrt{2\pi h^2}}e^{-\frac{(g_t - s_i)^2}{2h^2}} \tag{2}$$

where h is the bandwidth of Gaussian kernel function and it can be computed from the median absolute deviation (MAD)[8].

After background density estimation, the candidate moving objects can be extracted by subtracting the background image from each current frame. In the multimodal background model, background cannot be represented by a single image; therefore, the subtraction operation is implemented by probability thresholding. If the probability is less than a threshold T_p, the observed pixel probably belongs to the foreground, otherwise it is considered as background. The initial detection mask $\boldsymbol{M}_t$ is given by

$$\boldsymbol{M}_t(x, y) = \begin{cases} 1 & p_t < T_p \\ 0 & p_t \geqslant T_p \end{cases} \tag{3}$$

1.2 Noise removal

In outdoor surveillance systems, the camera is usually installed on an overbridge or high poles. When heavy vehicles pass the overbridge, the camera will experience small jitters and these introduce some noise in the detected foreground, particularly in many obvious edges. This noise is big in size. If exploiting the simple morphological operator to remove the noise, it will also incorrectly remove some small moving objects, e.g. people or vehicles near the road vanishing point. Due to the noise mainly resulting from the camera translation, we further classify the candidate foreground with the probability $p_{n,t}$ computed as

$$p_{n,t} = \max_{\bar{S} \in S_n}\left(\sum_{i=1}^{\bar{K}}\frac{1}{\sqrt{2\pi\bar{h}^2}}e^{-\frac{(g_t - \bar{s}_i)^2}{2\bar{h}^2}}\right), \boldsymbol{M}_t(x, y) = 1 \tag{4}$$

where $\bar{S}$ is one of the sample sets of the 8-neighbor of (x, y) and has $\bar{K}$ samples, S_n is a set of the 8-neighbor of S, $\bar{h}$ is the bandwidth of the 8-neighbor of (x, y) kernel function.

1.3 Shadow suppression

Color feature usually gives much information and is better than intensity in suppressing shadows. Due to the scalar color quantization of HMMD being equivalent to the vector color quantization of the hue-saturation-value (HSV) color space, we use the HMMD color space to detect shadows.

Although shadows share the same motion as the objects casting them, shadows have their particular

features. They have similar chromaticity but lower brightness than the corresponding background pixels. That is, a shadow cast on a background does not significantly change its hue and often lowers the saturation of the points[10].

For every pixel detected as candidate foreground, its corresponding samples are represented with hue, diff, and sum in the HMMD color space (see Fig.1) as $\boldsymbol{C}_i = (c_i^{\mathrm{h}}, c_i^{\mathrm{d}}, c_i^{\mathrm{s}})$. Given the observed pixel's color vector $\boldsymbol{C}_t = (c_t^{\mathrm{h}}, c_t^{\mathrm{d}}, c_t^{\mathrm{s}})$ at time t, the detected mask after shadow suppression can be computed by

$$\boldsymbol{M}_t' = \begin{cases} 0 & \boldsymbol{M}_t \cap \left(\left| c_{\mathrm{h},t} - \underset{i=1,2,\cdots,K}{\mathrm{med}} (c_{\mathrm{h},i}) \right| \leqslant T_{\mathrm{h}} \right) \cap \\ & \left(\left(c_{\mathrm{d},t} - \underset{i=1,2,\cdots,K}{\mathrm{med}} (c_{\mathrm{d},i}) \right) \leqslant T_{\mathrm{d}} \right) \cap \\ & \left(T_{\mathrm{s1}} \leqslant \dfrac{c_{\mathrm{s},t}}{\underset{i=1,2,\cdots,K}{\mathrm{med}} (c_{\mathrm{s},i})} \leqslant T_{\mathrm{s2}} \right) \\ \boldsymbol{M}_t & \text{otherwise} \end{cases} \tag{5}$$

where med(z) is the median function of samples. The use of median operator relies on an assumption that the background at every shadow pixel has many modals with little variation and will be visible more than fifty percent of the time during the training sequence. T_{s1} takes into account how strong the light source is. The stronger and higher the sun is, the lower value of T_{s1} is chosen. T_{s2} is less than one due to the shadow's being darker than the median value of the background. The choice of T_{d} and T_{h} is usually chosen empirically.

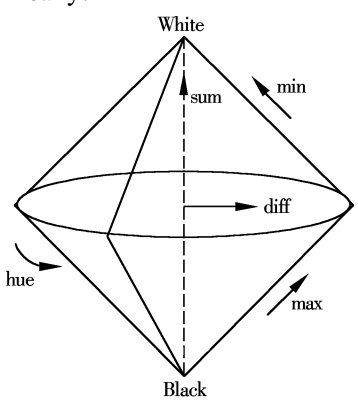

Fig.1 HMMD color space model

1.4 Model adaptation

In time the variation of scene should require background model changing. We update the model firstly by adding new n frames and disusing the oldest n frames. Then compute new inter-frame differences, and add the new background samples if the differences are less than T_{g}.

2 Experimental Results

The proposed algorithm is tested on a variety of outdoor traffic sequences. All sequences are captured with moving objects in the scene. In the first sequence, the sun is strong and high. Shadow is not main false positive, but the trees along the road wave with the wind load (see Fig.2).

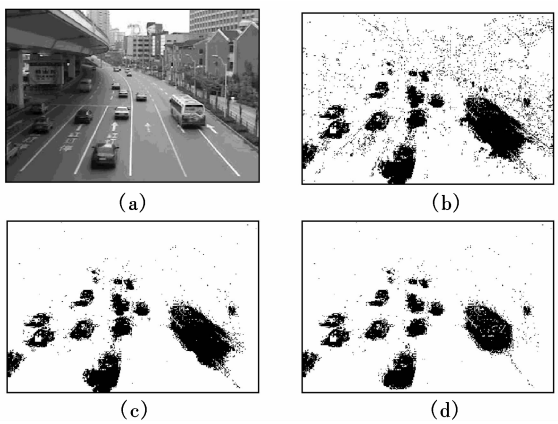

Fig.2 Detection results for sequence 1. (a) 6 218th frame (240 × 352); (b) Initial detection; (c) Noise removal; (d) Shadow suppression

The second sequence is captured at dusk, so shadows are longer and more obvious (see Fig.3).

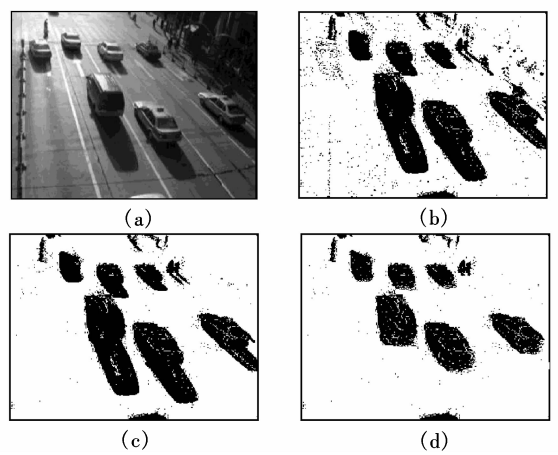

Fig.3 Detection results for sequence 2. (a) 210th frame (240 × 320); (b) Initial detection; (c) Noise removal; (d) Shadow suppression

Shadow suppression module aims to prevent moving cast shadows being misclassified as moving objects or parts of them, thus reducing the false positive. N is set to be 50 and $n = 2$; other parameters are set in Tab.2.

Tab.2 Parameters setting

Sequence	T_{g}	T_{p}	T_{s1}	T_{s2}	T_{d}	T_{h}
1	20	0.012	0.08	0.8	0.1	60
2	15	0.018	0.12	0.6	0.1	30

In Fig.2, there are many trees on the right side of the road waving with the wind. In initial detection results, there are many false positives due to the noise. After noise removal, most of them are removed. Image sequence in Fig.3 is captured at five to six o'clock in the afternoon. There are strong shadows that will cause serious problems without suppression. Note that the leftmost car is stationary over a long time, so it is **not detected as foreground but as background.**

3 Conclusion

We have presented a statistical multimodal background model for moving object detection in video sequences with both intensity and HMMD color information. The inter-frame difference is used to choose the coarse samples and this decreases the false negative. Noise and shadow suppression is processed to decrease the false positive. The proposed algorithm is tested in traffic sequences and can extract the **foreground with good performance.**

References

[1] McKenna S J, Jabri S, Duric Z, et al. Tracking groups of people [J]. *Computer Vision and Image Understanding*, **2000, 80**(1): 42 - 56.

[2] Rittscher J, Kato J, Joga S, et al. A probabilistic background model for tracking [A]. In: *Proceedings of the* 6*th European Conference on Computer Vision* [C]. Dublin, Ireland, 2000. 336 - 350.

[3] Wren C, Pentland A. Pfinder: real-time tracking of the human body [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **1997, 19**(7): 780 - 785.

[4] Friedman N, Russell S. Image segmentation in video sequences: a probabilistic approach [A]. In: *Proceedings of the* 13*th Conference on Uncertainty in Artificial Intelligence* [C]. San Francisco, CA: Morgan Kaufmann Publishers, 1997. 175 - 181.

[5] Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2000, 22**(8): 747 - 757.

[6] Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2000, 22**(8): 809 - 830.

[7] Toyama K, Krumm J, Brumitt B, et al. Wallflower: principles and practice of background maintenance [A]. In: *Proceedings of International Conference on Computer Vision* [C]. Kerkyra, Greece, 1999. 255 - 261.

[8] Elgammal A, Harwood D, Davis L S. Non-parametric model for background subtraction [A]. In: *Proceedings of the* 6*th European Conference on Computer Vision* [C]. Dublin, Ireland, 2000. 751 - 767.

[9] Manjunath B S, Ohm J R, Vasudevan V V, et al. Color and texture descriptors [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, **2001, 11**(6): 703 - 715.

[10] Prati A, Cucchiara R, Mikic I, et al. Analysis and detection of shadows in video streams: a comparative evaluation [A]. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* [C]. Kauai, Hawaii, 2001. 571 - 576.

基于噪声与阴影抑制多模态背景模型的运动物体检测

毛燕芬　　施鹏飞

(上海交通大学电子信息与电气工程学院, 上海 200030)

摘要: 针对场景照明变化、模型初始化以及阴影等问题, 提出了一种用于视频监视系统运动物体检测的统计多模态背景模型. 通过相隔固定的帧差值阈值化得到背景样本值, 并采用高斯核密度估计方法计算背景灰度的概率密度函数. 利用像素的邻域信息来去除由于摄像机抖动和场景小运动产生的噪声. HMMD 色彩信息用来检测和抑制运动投射阴影. 实验结果验证了算法在交通监控前景物体分割中的有效性.

关键词: 视频监视; 背景模型; 核密度估计; 阴影抑制; HMMD 色彩空间

中图分类号: TP391