# Video Image Segmentation Based on Bayesian Learning

WANG Lin-bo, ZHAO Jie-yu
(Research Institute of Computer Science and Technology, Ningbo University, Ningbo 315211)

**Abstract** Segmentation becomes a difficult task when the background illumination changes. In this paper, we apply a Bayesian learning method into video segmentation. The constantly changing background has been modeled at the pixel level. The feature vector for each pixel is represented with a discrete probability distribution function. The histogram colors and co-occurrence vectors have been calculated. Bayesian learning has been used to obtain these probability distribution functions from the video image inputs. The experimental results indicate that the proposed approach is able to learn a complex background of which the illumination changes either gradually or suddenly.

**Keywords** video image segmentation, bayesian learning, complex background modeling
**中图法分类号**: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2005)09-1073-06

# 基于贝叶斯学习的视频图像分割

王林波 赵杰煜
(宁波大学信息学院计算机科学与技术研究所,宁波 315211)

**摘 要** 如果背景中光线变化,那么视频图像分割将会变得比较困难。为了对光线变化的图像进行顺利侵害,提出了一种利用贝叶斯学习方法来进行视频图像分割的算法,即先在每个像素点处对不断变化的背景建模,同时计算每个像素点处的颜色直方图,再用这些直方图来表示该像素点处特征向量的概率分布,然后用贝叶斯学习方法来进行判断,以确定在光线缓慢或者突然变化的时候,每个像素点是属于前景还是属于背景。

**关键词** 视频图像分割 贝叶斯学习 复杂背景模型

## 1 Introduction

Video image segmentation is a process to detach the interesting objects and the background from the video clips. It is the first step towards target location and recognition. Various approaches have been developed for video image segmentation, such as the fast marching methods and region growing algorithms[1~3], and background modeling[4,5]. The former can't deal with problems caused by illumination changes, shadows, and homogeneous intensity on the moving objects. The background modeling approach organizes temporal information of sequence by virtue of sprite technique[4], it allows video sequence with various global motions, such as shifting, rotation and zoom. Other methods such as the skin-color based approach[6~10], which enables robust segmentation of skin-colored patches under time-varying illumination[6]. However, most of the existing techniques consider video segmentation problem only with a stationary background. If the light changes gradually, some

**基金项目**:国家自然科学基金项目(NSFC-60273094)
**收稿日期**:2004-11-09;**改回日期**:2005-03-29

**第一作者简介**:王林波(1979~ ),男。2005 年在宁波大学获硕士学位。主要研究方向为视频图像处理、模式识别。E-mail: wang_linbo@163.com

methods can still do video segmentation fairly well, but if the light changes suddenly, those algorithms will fail.

In this paper, we first model the background for both stationary points and changing points. Then we apply the Bayesian learning rule to obtain the probability distribution and segment the video image sequences with a gradually or suddenly changing background illumination.

## 2 Model Description

For most of the video segmentation problems, we are only interested in some foreground objects, and these objects are often moving, such as people, vehicles and so on. Except for these foreground objects, we consider other parts as background. The complex backgrounds may be caused by lighting variations, or some moving objects such as curtain in the wind, fountain, wavering tree branches etc, and these factors will severely influence the result of video segmentation.

There are two types of lighting variations, one is the gradual change caused by natural lighting variations; the other is the sudden "once-off" change caused by switching on or off some lights, which leads to obvious change to the environment. We establish a Bayesian model to learn the complex backgrounds. The model is used to classify every pixel as foreground or background.

### 2.1 Bayesian Model

Let $V_t$ be a discrete value feature vector extracted from an image sequence at pixel $s(x,y)$ and at time instant $t$. According to the Bayes rules, it follows that the posterior probability of $V_t$ from the background $b$ or foreground $f$ is

$$P(C|V_t,s)=\frac{P(V_t|C,s)P(C|s)}{P(V_t|s)},\ C=b\ \text{ or }\ f \quad (1)$$

Then the pixel is classified as background if the feature vector at this pixel satisfies:

$$P(b|V_t,s) > P(f|V_t,s) \quad (2)$$

Noting the feature vector $V_t$ at the pixel $s(x,y)$ is from background or foreground, it follows:

$$P(V_t|s) = P(V_t|b,s)\cdot P(b|s) + P(V_t|f,s)\cdot P(f|s) \quad (3)$$

Substituting (1) and (3) to (2), we can get:

$$2P(V_t|b,s)\cdot P(b|s) > P(V_t|s) \quad (4)$$

Therefore, as soon as we calculate the conditional probability of background $P(V_t|b,s)$, the priori probability of background $P(b|s)$ and the total probability $P(V_t|s)$ at pixel $s(x,y)$, we will know this pixel is of background or foreground.

### 2.2 Representation of Feature Vector

It is important to appropriately represent the feature vector $V_t$ extracted from an image sequence at the pixel $s(x,y)$ and time instant $t$. If the background is stationary at a point, we represent the feature vector $V_t$ as $[r_t,g_t,b_t]^T$, which is just the color value at pixel $s(x,y)$ and time instant $t$. If the background is changing, we represent the feature vector $V_t$ as $[r_{t-1},g_{t-1},b_{t-1},r_t,g_t,b_t]^T$. The color value of one pixel in a color image is of three dimensions, and each dimension as 256 quantization levels, operating on the joint histogram would be expensive both for computation and storage. So a good approximation is desirable to decrease the quantization levels.

### 2.3 Representation of Probability

In the inequality (4), the conditional probability $P(V_t|b,s)$ and the total probability $P(V_t|s)$ are usually unknown, they could be represented by the histograms of feature vectors over the entire feature space.

In general, if the background is stationary or the light changes gradually or suddenly, the feature vector at one pixel of the background would concentrate to a very small subspace of the feature histogram, while the feature vector would distribute widely in the feature space when the moving objects pass on it.

We suppose there are $M$ different feature vectors at the pixel $s(x,y)$, and every feature vector is sorted according to the descending order of $P(V_t|b,s)$. Then there must be a number $N(N<M)$, which satisfied the next inequalities:

$$\sum_{i=1}^{N} P(v_t^{(i)}|b,s) > 0.95 \text{ and } \sum_{i=1}^{N} P(v_t^{(i)}|f,s) < 1 - 0.95 \quad (5)$$

We can consider the first $N$ feature vectors as the

significant portions of background at this pixel.

In order to classify the pixel, we design a table of probability, whose element consists of

$$S_{V_t}^{(s,t,i)}=\begin{cases}p_{\mathrm{v}}^{(s,t,i)}=P(\boldsymbol{v}_t^{(i)}\mid s)\\ p_{\mathrm{vb}}^{(s,t,i)}=P(\boldsymbol{v}_t^{(i)}\mid b,s)\\ \boldsymbol{v}_t^{(i)}=[a_1^{(i)},\cdots,a_n^{(i)}]^{\mathrm{T}}\end{cases}\tag{6}$$

This table will be used in the following algorithm.

## 3 Algorithm Description

The algorithm consists of two main parts. The first part is the batch learning of the initial background images. The second part is the implementation of video image segmentation and continuous learning for the changing background. The main blocks of the system are given in Fig. 1.

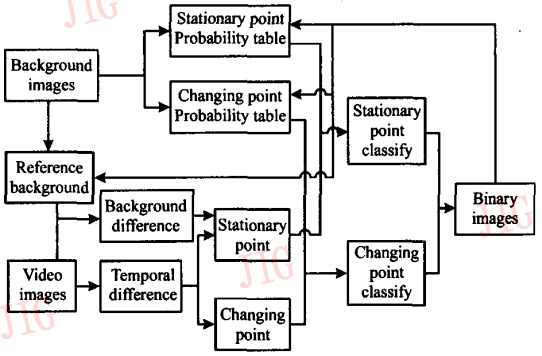



Fig. 1　Algorithm Flow

The detailed cycles of the algorithm are as follows:

(1) Learn the probability distribution function and feature vectors for each pixel from the initial background.

(2) Compute the temporal difference $F_{\mathrm{td}}(s,t)$ between the current two successive input frames, and the background difference $F_{\mathrm{bd}}(s,t)$ between the current frame and the reference background.

(3) Compare each element of $F_{\mathrm{td}}(s,t)$ with threshold $T$, if $F_{\mathrm{td}}(s,t)>T$, we consider the pixel $s(x,y)$ as a changing point, otherwise as a stationary point.

(4) For all the stationary and changing pixels, calculate the probability $P(b\mid s)$, $P(\boldsymbol{V}_t\mid s)$ and $P(\boldsymbol{V}_t\mid b,s)$ according to

$$\begin{cases}P(b\mid s)=p_{\mathrm{b}}^{(s,t)}=\sum\limits_{j=1}^{N}p_{\mathrm{v}}^{(s,t,j)}\\ P(\boldsymbol{V}_t\mid s)=\sum\limits_{j\in M(\boldsymbol{V}_t)}p_{\mathrm{v}}^{(s,t,j)}\\ P(\boldsymbol{V}_t\mid b,s)=\sum\limits_{j\in M(\boldsymbol{V}_t)}p_{\mathrm{vb}}^{(s,t,j)}\end{cases}\tag{7}$$

Where, $p_{\mathrm{v}}^{(s,t,j)}$ is the total probability of the $j$th feature vector of the probability table at pixel $s(x,y)$ and time instant $t$. $p_{\mathrm{vb}}^{(s,t,j)}$ is the $j$th feature vector's conditional probability of background in the probability table at pixel $s(x,y)$ and time instant $t$. $\boldsymbol{M}(\boldsymbol{V}_t)$ is defined as the matched feature set: $\boldsymbol{M}(\boldsymbol{V}_t)=\{i:|\boldsymbol{v}_t-\boldsymbol{v}_t^{(s,i)}|<\delta\}$, $i\in[1,N]$ and the variable $\boldsymbol{v}_t^{(s,i)}$ is the $i$th feature vector of the probability table at pixel $s(x,y)$, $\delta$ is a match threshold.

(5) If the inequality $2P(\boldsymbol{V}_t\mid b,s)\cdot P(b\mid s)>P(\boldsymbol{V}_t\mid s)$ holds, consider the pixel as one of background, set the color value to 0 at pixel $s(x,y)$; otherwise, set the value to 1. At the end of this step, a binary image is generated.

(6) A morphological operation (open) is used to remove the scattered noise points. At last, the system outputs a video segmentation image.

(7) Carry on the learning process by the continuous updating of the probability tables and the reference background in order to adapt the system to the changing environment. The details of continuous learning process are described in the next subsection.

(8) Go back to step (2).

During the initial learning process (step 1), all kinds of background images are learned by the system, and these background images may include some complex objects. At the end of the initial learning process, the system would obtain two probability tables including $P(\boldsymbol{V}_t\mid b,s)$, $P(\boldsymbol{V}_t\mid s)$, and feature vectors at every pixel. Here we set $P(\boldsymbol{V}_t\mid b,s)=P(\boldsymbol{V}_t\mid s)$ because all input images belong to the background. The system also gets a reference background which represents the most recent appearance of the background. This reference background is maintained at each time step during the continuous learning process to make the background difference accurate. The background differencing is

used to remove the image noise.

**Updating the Probability Table**

The tables of probability are maintained at each pixel. Two updating strategies are proposed to both gradual and sudden illumination changes. When the lighting changes gradually in the environment, the updating rules are:

$$\begin{cases} p_{\mathrm{v}}^{(s,t+1,i)} = (1-\alpha)p_{\mathrm{v}}^{(s,t,i)} + \alpha M_{\mathrm{v}}^{(s,t,i)} \\ p_{\mathrm{vb}}^{(s,t+1,i)} = (1-\alpha)p_{\mathrm{vb}}^{(s,t,i)} + \alpha (M_{\mathrm{b}}^{(s,t)} \wedge M_{\mathrm{v}}^{(s,t,i)}) \end{cases} \quad \alpha \in [0,1] \tag{8}$$

where, $\alpha$ is a learning rate. Given $\boldsymbol{v}_t$ as an input vector at pixel $s(x,y)$, and $\boldsymbol{v}_t^{(s,i)}$ as a feature vector in the probability table matching $\boldsymbol{v}_t$ best,

$$M_{\mathrm{v}}^{(s,t,i)} = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$M_{\mathrm{b}}^{(s,t)} = 1$ when the pixel at $s(x,y)$ is labeled as background at time instant $t$ from the feedback of final segmentation, otherwise $M_{\mathrm{b}}^{(s,t)} = 0$. The symbol '$\wedge$' is a logic AND operation.

The second function in (8) states that the appearance probability of the feature vector $\boldsymbol{v}_t^{(s,i)}$ matching best in the probability table is increased due to $M_{\mathrm{v}}^{(s,t,i)} = 1$. If $M_{\mathrm{v}}^{(s,t,i)} = 0$, the appearance probability of the feature vectors in the probability table are slightly decreased. If the input vector $\boldsymbol{v}_t$ does not match any feature vectors, it will replace the last element in the probability table as a new feature vector, and a small number is assigned to the probability of new feature vector.

When the lighting changes suddenly in the environment, the new features of the new background appearance become dominated immediately. From (3) and (5), the new background feature vectors are detected if

$$P(f \mid s) \sum_{i=1}^{N} P(\boldsymbol{v}_t^{(i)} \mid f,s) > 0.95 \tag{10}$$

And from (3), (6), we can conclude that

$$\sum_{i=1}^{N} p_{\mathrm{vb}}^{(s,t,i)} \cdot \left(1 - \sum_{i=1}^{N} p_{\mathrm{v}}^{(s,t,i)}\right) > 0.95 \tag{11}$$

Noting that, the updating functions are

$$\begin{cases} p_{\mathrm{v}}^{(s,t+1,i)} = 1 - p_{\mathrm{v}}^{(s,t,i)} \\ p_{\mathrm{vb}}^{(s,t+1,i)} = (p_{\mathrm{v}}^{(s,t,i)} - p_{\mathrm{b}}^{(s,t)} \cdot p_{\mathrm{vb}}^{(s,t,i)}) / p_{\mathrm{b}}^{(s,t+1)} \end{cases} \quad i \in [1,N] \tag{12}$$

where, $p_{\mathrm{b}}^{(s,t)} = \sum_{i=1}^{N} p_{\mathrm{v}}^{(s,t,i)}$, $p_{\mathrm{b}}^{(s,t+1)} = \sum_{i=1}^{N} p_{\mathrm{v}}^{(s,t+1,i)}$

**Updating Reference Background**

Reference background images represent the most recent appearance of the background in the video. If the background changes gradually, the reference background is updated as

$$B_{\mathrm{c}}(s,t+1) = (1-\beta)B_{\mathrm{c}}(s,t) + \beta I_{\mathrm{c}}(s,t) \quad \beta \in [0,1] \tag{13}$$

where, $\beta$ is an update rate of background, and $B_{\mathrm{c}}(s,t)$ is the color value at pixel $s(x,y)$ in the reference background, and $I_{\mathrm{c}}(s,t)$ is the color value at pixel $s(x,y)$ in the current frame. If $F_{\mathrm{id}}(s,t) = 1$ and the color value at one pixel in the video segmentation image is 0, or $F_{\mathrm{bd}}(s,t) = 1$, both show that the background has changed a lot, so the reference background is updated as

$$B_{\mathrm{c}}(s,t+1) = I_{\mathrm{c}}(s,t) \tag{14}$$

## 4 Experimental Results

We use some background video clips for the initial learning of the system. Two different histograms at every pixel of each frame have been calculated. Thinking of storage and computation, the quantization levels of feature vectors should be decreased. There are only three dimensions in $\boldsymbol{V}_t = [r_t, g_t, b_t]^{\mathrm{T}}$, so $\boldsymbol{V}_t = [r_t, g_t, b_t]^{\mathrm{T}}$ is decreased from 256 to 64, and $64^3$ bins in the joint histogram for the whole feature vector space; while $\boldsymbol{V}_t = [r_{t-1}, g_{t-1}, b_{t-1}, r_t, q_t, b_t]^{\mathrm{T}}$ is decreased from 256 to 32, and $32^3$ bins in the joint histogram for the whole feature vector space.

The system calculates two probability tables of every pixel of each frame according to the histogram generated. One is a probability table for stationary points, and the other is a probability table for changing points. In the probability table for stationary points, if the number of feature vectors at pixel $s(x,y)$ is more than 50, the first 50 feature vectors are stored according to the descending order of $P(\boldsymbol{V}_t \mid b,s)$, otherwise all feature vectors are stored. In the probability table for changing points, if the number of feature vectors at pixel $s(x,y)$ is more than 80, the

first 80 feature vectors are stored according to the descending order of $P(V_t \mid b,s)$, otherwise all feature vectors are stored.

The last frame in the initial background image sequence is chosen to be the reference background.

Variables such as the learning rate $\alpha$, the update rate $\beta$ of background, and the match threshold $\delta$, are initialized before the segmentation. For the learning rate $\alpha$, if it is too small, the system will become too slow to response the sudden background changes. In the experiment, set the learning rate to be a fixed value from 0.005 to 0.01. The update rate $\beta$ of background controls the updating rate of the reference background. If the update rate $\beta$ of background is too large, foreground will become the background at once; and if it is too small, the reference background can not reflect the current background well. We set update rate $\beta$ of background at 0.05 in the experiment. The match threshold $\delta$ is set to 30.

Two experiments have been performed on different scenes where the lighting changes gradually and suddenly.

In the first experiment, the most significant features of background are learned when we drew the curtain slowly to make the lighting indoor change gradually, and after that, a person moved in front of a camera with the light changing gradually, The results are shown in Fig. 2.

(a)

(b)

Fig. 2　Segmentation results with gradually changing light

The images in the sequence (a) are extracted from the video with an interval of 5 frames, and images in the sequence (b) are the segmentation results of the corresponding frames. It can be seen that the proposed Bayesian learning method works very well when the lighting changes gradually.

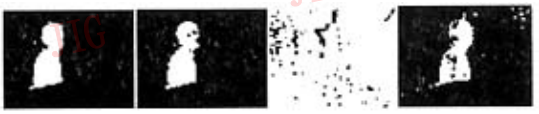
In the second experiment, the most significant features of background are learned when we switched the lights on or off to have a sudden change of the illumination. And then, a person walked in front of a camera with the lights on or off to have a sudden change of the illumination, The results of video segmentation are shown in Fig. 3 and Fig. 4:

(a)

(b)

Fig. 3　Segmentation results when light turns off

(a)



(b)

Fig. 4　Segmentation results when light turns on

In Fig. 3, images in the sequence (a) are continuous video frames when the lights are switched off, and images in the sequence (b) are the segmentation results.

In Fig. 4, images in the sequence (a) are the continuous video frames when the lights are switched on, and (b) is the result of video segmentation.

From Fig. 3 and Fig. 4, we can find that the result

of video segmentation based on Bayesian learning is good whenever the lighting changes from bright to dark or from dark to bright, and the adaptive capability is strong to the condition with suddenly changed illumination.

## 5 Conclusions

We use a Bayesian learning method to enhance the video segmentation when the lighting changes gradually or suddenly in the background. The system initially computes the histogram of color or co-occurrence vectors at every pixel, and generates two probability tables, in which the feature vectors and the probability distribution function are included. All these feature vectors represent the background at every pixel of each frame. Then, a Bayesian method is used to judge if every pixel belongs to either background or foreground. Two probability tables are also learned online to adapt to the changing background.

Future work is to increase the robustness of the proposed method, and apply it in more complex backgrounds including some moving objects. Based on that, we will do target location and object recognition.

### References


1 Sifakis E, Grinias I, Tziritas G. Video segmentation using fast marching and region growing algorithms[J]. EURASIP Journal on Applied Signal Processing, 2002,**4**: 379 ~ 388.

2 Sifakis E, Tziritas G. Fast marching to moving object location[A]. In: Proceedings 2nd International Conference on Scale-Space Theories in Computer Vision[C], Corfu, Greece,1999: 447 ~ 452.

3 Grinias I, Tziritas G. A semi-automatic seeded region growing algorithm for video object localization and tracking[J]. Signal Processing: Image Communication, 2001,**16**(10): 977 ~ 986.

4 Lu Y, Gao W, Wu F. Automatic video segmentation using a novel background model[A]. In: IEEE International Symposium on Circuits and Systems [C], Scottsdale, Arizona, USA, 2002:808 ~ 810.

5 Li L Y, Huang W M, Gu Y H, *et al*. Foreground object detection from videos containing complex background[A]. In: Proceedings ACM Multimedia Conference [C], Berkeley, California, USA, 2003: 2 ~ 10.

6 Sigal L, Sclaroff S, Athitsos V. Skin color-based video segmentation under time-varying illumination[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004,**26**(7): 862 ~ 877.

7 Birchfield S T. Elliptical head tracking using intensity gradients and color histograms[A]. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition [C], Santa Barbara, California, USA, 1998: 232 ~ 237.

8 Hafner W, Munkelt O. Using color for detecting persons in image sequences[J]. Pattern Recognition and Image Analysis, 1997, **7**(1):47 ~ 52.

9 Yang J, Weier L, Waibel A. Skin-color modeling and adaptation [A]. In: Proceedings of the 3rd Asian Conference on Computer Vision[C], Hong Kong, 1998:687 ~ 694.

10 Storring M, Andersen H J, Granum E. Skin colour detection under changing lighting conditions[A]. In: Proceedings 7th Symposium on Intelligent Robotics Systems [C], Coimbra, Portugal, 1999:187 ~ 195.