ABSTRACT

Segmentation of an interesting target from a dynamic video has been an important research topic in computer vision. In this work, we present a novel recursive Bayesian learning method for dynamic video segmentation. In the algorithm, each frame pixel is represented as layered normal distributions and the recursive Bayesian estimation is used to update the background parameters so as to obtain a robust background model. In the segmentation, foreground is separated by simple background subtraction method firstly. And then, a local texture correlation operator is proposed to remove vacancies in the separated foreground to refine the segmentation result. Experiments with two typical video clips are used to demonstrate that the proposed method can outperform traditional methods in both segmentation result and converging speed.

Index Terms— Image segmentation, video processing, recursive estimation, Bayesian learning

1. INTRODUCTION

Video segmentation for interesting target detection and tracking has been an important research issue in intelligent surveillance and many other vision systems. A lot of segmentation algorithms have been proposed in previous works for different applications, such as Gaussian mixture model (GMM) [1], non-parametric kernel density estimation (KDE), topology free Hidden Markov Model (HMM), Kalman filter (KF), and Bayesian background model (BBM) [2] etc. In these methods, statistical model is usually used to represent the pixel-wise variation of visual features along the temporal domain. Performance of these algorithms greatly depends on the modeling of the background and its update.

GMM is one classical means for the changing background modeling, such as the swaying trees, water movement or ambient light changes. The model parameters for each Gaussian model are updated using an online Expectation Maximization (EM) to adapt the background changes [3-5]. In [5], a nonparametric kernel density estimation method for background modeling is proposed. A kernel-based function is employed to represent the color distribution of each background pixel. The kernel based distribution is a generalization of GMM which requires no parameter estimation. The method has two significant advantages vs. original GMM. First, the method has no assumptions to the underlying distributions. Second, the method needs not to determine the number of distributions. Generally, GMM based methods are usually subject to its huge computation and low converging velocity and that makes them impractical to real-time segmentation tasks. Except for GMM, there are also a lot of other algorithms have been proposed. In [6], the distribution of temporal variations in color at each pixel is used to model the spectral feature of the background as well as its update. In [7], the motion information is used to model the dynamic scenes. HMM is an effective method to represent the scene in discrete states corresponding to environmental conditions and switching among these states with the observations [8]. In [9], a topology free HMM algorithm is proposed to process the sudden changes of illumination in the background modeling application. Moreover, Kalman filter has also been used for real-time tracking of illumination changes in the background model [10]. But, the method cannot deal with the changes of objects in the background, which usually causes incorrect segmentations. In addition, Wiener Filter and Conditional Random Field are also applied in the pixel based background modeling. To improve the segmentation efficiency, some image block based dynamic background models are also presented. In [11], the spatial concurrence phenomenon between image block changes is used to realize the foreground segmentation. In [12], the normalized vector distance is introduced to measure the correlation between blocks and improved the segmentation result. In [13], a local binary pattern (LBP) histogram is used to describe the background model. The main disadvantage of block based background modeling methods is that the segmented objects are lack of accuracy.

In this paper, the background model is firstly initialized with layers of Gaussian distributions for each frame pixel. To improve the calculation efficiency, only pixels satisfy a given confidence value are considered as foreground pixels. And then, recursive Bayesian estimation is used to update...
the background parameters so as to achieve robust background segmentation. To deal with the vacancy phenomenon in the primary segmentation, a local texture correlation operator is proposed.

The paper is organized as follows. Section 2 describes the proposed algorithm. Experimental results on two real video clips and the comparison with traditional algorithms are presented in Section 3. Conclusion and future work can be found in Section 4.

2. PROPOSED METHOD

2.1 Background Modelling and Update

To each frame, its image pixel is represented with layers of 3-dimension multivariate Gaussian function. And each layer corresponds to a different appearance of the pixel. Assume that the pixels all follow multivariate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). According to Bayesian decision theory, the \( p(\mu, \Sigma) \) can be written as:

\[
p(\mu, \Sigma) = p(\mu|\Sigma)p(\Sigma)
\]  

Multivariate statistical analysis is built on multivariate normal distribution and samples covariance matrix based on multivariate normal distribution obeys Wishart distribution. Define the joint prior distribution of covariance matrix to normal inverse-Wishart distribution, the mean and covariance matrix are assumed to be unknown. We have following equation:

\[
\Sigma 
\sim \text{W}^{-1}(\mathbf{C}_{n-1}, N - 1)
\]  

where \( \mathbf{C} \) indicates scale matrix for Normal-Inverse-Wishart (NIW) distribution, and \( N - 1 \) is the degree of freedom. The probability density function of \( \Sigma \) can be expressed as:

\[
f(\Sigma; \varphi_n, \mathbf{C}) \propto |\mathbf{C}|^{n+N-1/2} |\Sigma|^{-(n+N)/2} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{C}_{n-1})}
\]

where \( \text{tr}(\cdot) \) represents the matrix trace. With known distribution of \( \Sigma \), \( \mu \) obeys the law of multivariate normal distribution:

\[
\mu \sim \text{N}_{n}(\mu_{n-1}, \xi_{n-1} \Sigma)
\]

Therefore, the joint prior distribution of \( (\mu, \Sigma) \) following NIW distribution can be expressed as:

\[
f(\mu, \Sigma | x) \propto |\Sigma|^{-(n+N)/2} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{C}_{n-1})} \times \frac{1}{2} (\mu - \mu_{n-1})(\xi_{n-1} \Sigma^{-1}(\mu - \mu_{n-1}) + \text{tr}(\Sigma^{-1} \mathbf{C}_{n-1}))
\]

where \( \xi_{n-1} \) indicates the number of prior measuring parameters. We first label the joint density \( f(\mu, \Sigma) \) as \( \text{NIW}(\mu_{n-1}, \varphi_{n-1}, \mathbf{C}_{n-1}) \). By using Bayesian statistics theory and multiplying prior density with the normal likelihood, the joint posterior density \( \text{NIW}(\mu, \varphi, \mathbf{C}) \) with the parameters updated recursively can be formulated as following equations:

\[
\mu_n = \frac{\mu_{n-1} + 1}{1 + \varphi_{n-1}} + \frac{\varphi_{n-1}}{1 + \varphi_{n-1}} (\bar{x} - \mu_{n-1})
\]

\[
\varphi_n = \varphi_{n-1} + \beta
\]

\[
\mathbf{C}_n = \frac{\mathbf{C}_{n-1} + \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T}{\varphi_n} + \beta (\bar{x} - \mu_{n-1})(\bar{x} - \mu_{n-1})^T
\]

where \( \bar{x} \) indicates the mean of new samples and \( \beta \) is the number of samples used to update the model, \( \xi_n \) is the number of prior measuring parameters after the update. Confidence decision \( D \) for the layer is defined as:

\[
D = (\varphi_n - 2)^4 / (\xi_n (\varphi_n - 4) |\mathbf{C}_n|)
\]

To update the background model, it is firstly initialized to \( \varepsilon \) layers for each pixel. The number of layers normally varies with the number of dynamic scenes. The update algorithm starts from the most confident layer in the model. If the observed sample locates in \( \lambda \) confidence interval of the current model, model parameters are updated via Eqn. (7)-(10). If current sample is outside the confidence interval, the number of prior measuring parameter is updated as:

\[
\xi_n = \xi_{n-1} / (1 + \beta \xi_{n-1})
\]

Once none of the models are updated, the least confident layer will be removed. Flow chart of the update algorithm for individual pixel is as shown in Fig. 1.

2.2 Foreground Segmentation and Refinement

The preliminary foreground is obtained by subtracting the segmented background. However the vacancy phenomenon becomes seriously when the foreground and background are of the similar color or grey level. To improve the final foreground segmentation result, a local texture correlation method is proposed in this work. It has been proved that, pixels in a relative small image region between background and foreground have local texture correlation property [10]. Pixels in each neighbourhood are reclassified according to the above cross correlations to compensate small holes within foreground objects. Assuming the grey level at image
position $p$ in the $t$-th frame is $f(x,y)=f(x,y|(t,p))$. Then gradient vector $G$ of this point can be described as:

$$G(f(x,y|(t,p))) = \nabla f(x,y|(t,p)) \vec{j} + \nabla f(x,y|(t,p)) \vec{j}$$  \hspace{1cm} (12)

Given two adjacent pixels $p_i, p_j$ in a smaller neighborhood of $p$, then gradient vector similarity $S$ for $p_i$ and $p_j$ is:

$$S(p_i, p_j) = \frac{G(f(x_i, y_i|(t,p_i))) \cdot G(f(x_j, y_j|(t,p_j)))}{\|G(f(x_i, y_i|(t,p_i)))\| \cdot \|G(f(x_j, y_j|(t,p_j)))\|}$$  \hspace{1cm} (13)

By the similarity definition, we have:

$$S(p_i, p_j) + S(p_j, p_i) \geq 2S(p_i, p_j)$$  \hspace{1cm} (14)

It shows that if $p_i$ and $p_j$ belong to either foreground or background, the local texture feature of $p_i$ and $p_j$ is similar. Therefore:

$$S(p_i, p_j) + S(p_j, p_i) \approx 2S(p_i, p_j)$$  \hspace{1cm} (15)

On the other hand, if $p_i$ and $p_j$ come from the surface of different objects, the local texture feature of $p_i$ and $p_j$ usually have big difference. The criterion above can be realized by introducing the discriminate function $\omega(p_i, p_j)$:

$$\omega(p_i, p_j) = 1 - \frac{2S(p_i, p_j)}{S(p_i, p_j) + S(p_j, p_j)}$$  \hspace{1cm} (16)

The threshold of $\omega$ is usually set to 0.2 experientially.

The algorithm starts with following initial parameters:

$$\xi_0 = 0.1, \phi_0 = 10, \mu_0 = z, \epsilon = 5, \lambda \in (0.9,1.0), C_0 = (\phi_0 - 4)I,$$

where $I$ indicates the three dimensional identity matrix. The experiments are conducted on two typical video clips which contain some changing background like the swaying trees and floating highway. The system is running on a P4-2GHz desktop with 1GB RAM. The algorithm is also compared with GMM and BBM and the results are as shown in Fig. 2 and Fig. 3. In Fig.2 and Fig.3, the left column is the original video frames: 326, 352 and 405 (415, 439 and 488), the 2nd, 3rd and 4th rows displayed the background segmentation results by GMM, BBM and Our method respectively. In comparison, the proposed method outperforms in dynamic scenes (swaying trees in Fig.2 and floating highway in Fig.3) and also gives better segmentation results without jamming of background noise and the vacancy phenomenon that exist in GMM and BBM methods.

In addition, the stability of background changes with respect to the frame is provided in Fig. 4. From the comparison we can see that the extracted background by our method reaches to the stable state (around 400th frames) faster than GMM and BBM.
4. CONCLUSION AND FUTURE WORK

This paper presents a novel recursive Bayesian learning method for video segmentation with dynamic background. In the algorithm, each frame pixel is represented as layered normal distributions and the recursive Bayesian estimation is used to update the background parameters so as to obtain a robust background model. In the segmentation, foreground is separated by simple background subtraction method firstly. And then, a local texture correlation method is introduced to remove vacancies in the separated foreground to achieve better segmentation result. Experimental results on two typical video clips are demonstrated to show the proposed method outperforms previous methods like GMM and BBM in both segmentation result and converging speed.

Future work can address how to deal with more challenging scenarios and how to improve algorithm converging and the system running speed further.

5. ACKNOWLEDGMENT

The work is partially supported by Knowledge Innovation Program of the Chinese Academy of Sciences (Grant no. KG032-YW-156) and the NSFC (Project no. 60903115).

6. REFERENCES