Image Object Detection Algorithm Based on Improved Gaussian Mixture Model

LI Xing-liang  
Department of Criminal Science and Technology, Nanjing Forestpolice College, Nanjing Jiangsu, 210023, China  
Email: ebeauty_lee@126.com

WU Yu-bao  
Department of Information Technology, Nanjing Forestpolice College, Nanjing Jiangsu, 210023, China  
Email: eice_wu@126.com

Abstract—Aiming at poor adaptability to illumination variation and single learning rate in traditional Gaussian mixture model, an improved moving object detection algorithm based on adaptive Gaussian mixture model is proposed in this paper, so as to achieve the goal of a self-adaptive background updating model. In this paper, we analyze the existed algorithms and put forward the method to make use of color histogram matching algorithm, through introduction of illumination variation factor and update-counter of model parameter and the components number with self-adaptive selection employed to adaptively adjust learning rate, in order to greatly reduce the computation time of the algorithm and improve the real-time performance. The experiment results show that the new method can effectively adapt the scene, and has more good expansibility, robustness and stability than traditional Gaussian mixture model.

Index Terms— Gaussian Mixture Model; Adaptive; Moving Target Detection; Histogram Matching

I. INTRODUCTION

The so-called moving target detection means to extract target from streaming video in real-time by determining the target’s region and color feature. The result of it is static as the foreground target described by some static features. According to the background environment of the moving target, there are two types of detection: moving target detection with static background and with dynamic background. Relative to still information, motion information can attract people’s attention more and also is more useful. To illustrate, for automatic video surveillance system, the change of monitoring object is one of the information that correspond with the system well. The system utilizes those image sequences collected by camera to monitor its target uninterruptedly. If abnormal condition takes place, it gives alarm signal to notify related people for doing it. For detecting the change of monitoring objects, the system has to handle those image sequences effectively in order to isolate dynamic information, the target’s motion, in usual. When we convert those detected images into digital images which can be processed by computer, the quality of images will be lower disappointedly and some noises will be caused by several reasons during the whole conversion. Those noises can influence the post-processing of those images a lot. So in every image processing system, they always add picture pre-processing into the system for making images more suitable for image subsequent processing.

Moving target’s detection belongs to the low-level stage process of computer vision. It’s also the basic process of achieving target tracking, feature extracting, and behavior analyzing and understanding further. Therefore, effective detection results are vitally crucial for the next treatment to high-level stage process. To demonstrate, the methods of this detection include point detection, image segmentation, inter-frame difference, background subtraction, cluster analysis and the motion vector field analysis. Among them, background subtraction method is widely applied to the process especially in the scene of fixed camera. It separates background and foreground by comparing the background model updated in real-time from every single frame of a moving target.

Gaussian Mixture Model (GMM) is a kind of background modeling method based on pixel level. Reference [3] puts it into use of background modeling method based on pixel level. Reference [3] puts it into use of background modeling method based on pixel level. Reference [4] separates background learning with background updating based on original GMM into two stages. The learning rate of background initializing stage is 1/N. Then, it adopts normal iterative manner to update. Literature [5] proposes an adaptive method that is each Gaussian distribution with different learning rate, and this method improves convergence speed. Reference [6] analyzes the phenomenon of “pixel saturation” caused by the serious deterioration of the convergence of variance when multimodal means and variance update, and it suggests two different learning factors that update means and variances separately. In some recent researches, reference [7] proposes a window weight update method in order to reduce the system runtime and improve the real-time character. Through various scholars’ studies, the robustness and stability of GMM have a certain extent
enhancement; However, the suitability to background changes is not as satisfying as them especially to some phenomenon like sudden illumination change. Reference [8] proposes to judge whether sudden illumination change happens through counting the amount of foreground pixels and the whole image pixels and comparing the given ratio. At the same time, the current changed image frames cannot be updated effectively and timely.

Based on traditional GMM, illumination variation factor is introduced to solve the low dynamic adaptation capability problem caused by illumination variation. Through adaptive update of learning rate and meanwhile the components number with self-adaptive selection which can reduce the computation time, the robustness and stability of model have a certain extent enhancement.

II. THE TARGET IMAGE PREPROCESSING

The image’s color can be divided into two kinds: black-and-white, and color. Black and white means that there are not any other colors in the image, only including black and white. In the RGB color model, if R=G=B, then the color (R, G, B) means one kind of black and white and the value of R=G=B is called gray value. The process of turning color into gray can be called gray processing. The graying is the process to make the color R, G, B’s quantity equal. For the value range of R, G, B is 0~255, the gray level only has 256 levels, that is to say the graying image can only express 256 colors.

There are three methods of gray processing in the following:

(1) Maximum value method: we can make the value of R, G, B equal to the largest value of the three values, namely,

\[ R = G = B = \max(R, G, B) \]  

(1)

The maximum value method can form the gray image with high brightness.

(2) Mean value method: we can get the mean value of R, G, B, namely,

\[ R = G = B = (R, G, B)/3 \]  

(2)

The mean value method can form the gray image with relative softness.

(3) Weighted mean value method: we can assign R, G, B different weighted values according to its importance or other indexes and make the R, G, B’s values are weighted, namely,

\[ R = G = B = (W_R R + W_G G + W_B B)/(W_R + W_G + W_B) \]  

(3)

where, \( W_R, W_G, W_B \) is R, G, B’s weight value respectively. \( W_R \) can be given different values and the weighted mean value method can form different gray images. Because human eyes have high sensitivity to green color, lower sensitivity to red color and the lowest sensitivity to blue color, so we can get a relative reasonable gray image to make \( W_G > W_R > W_B \). The experiment and theory derivation prove that, when \( W_R = 0.30, W_G = 0.59, W_B = 0.11 \), namely,

\[ R = G = B = 0.3R + 0.59G + 0.11B \]  

(4)

At this time, the R, G, B’s values are the pixel’s brightness, and we can get the most reasonable gray image now. We use the weighted mean value method to carry on the collected original image’s gray processing. The gray processing result and the image smoothing filter can be divided into linear filter and nonlinear filter. According to the linear filter, we mainly adopt the mean filter.

Mean filter is that we compute in the image’s local mean value and each pixel value can be replaced by the mean value of all values in its local neighborhood domain, that is:

\[ c_{i\alpha\rho} \]  

(5)

where, \( f(m, n) \) is a \( M \times N \) road image in size, and the mean filter window is \((2K+1)(2L+1)\), which must be odd numbers in two directions, or the image will produce deflection.

The theory and experiment prove that, though the linear filter has fine noise reduction performance, its image smoothing will cause the image’s detail information loss, which make the processed image fuzzy.

III. THE BACKGROUND MODELING BASED ON GMM

As for the simple scene, we can use the single Gaussian model to express each pixel’s color vector change. However, due to the complicated scene, the single mode’s distribution cannot get the effective fitting data. GMM is the weighted sum of a finite number Gaussian functions which can describe the pixel’s multi-peak state and conduct the complicated and dynamic background modeling.

The traditional GMM supposes that each pixel point in images is mutually independent, and the changes in time domain can be simulated by K multi-dimensional Gaussian distribution. The sample value of a certain pixel point \( P(x, y) \) is \((x, x_1, ..., x_J)\), the probability that we can observe the present pixel value \( x \) in t time is:

\[ P(X_i) = \frac{1}{\alpha \iota} \sum_{i=1}^{K} \sigma_{i\alpha\rho} N(X_i, \mu_{i\alpha\rho}, \sum_{i\alpha\rho}) \]  

(6)

where, \( K \) is the quantity of model components, and \( \omega_{i\alpha\rho} \), \( \mu_{i\alpha\rho} \) and \( \sum_{i\alpha\rho} \) is the weighted value, mean value and covariance matrix of the i gaussian distribution of the model in t time. The probability density function of the i gaussian distribution is

\[ N(X_i, \mu_{\iota\sum}) = \frac{1}{(2\pi)^{i/2}} |\sum_i^{\alpha\rho}|^{-1/2} \exp(\frac{1}{2}(X - \mu_{\iota\sum})^T \sum_i^{\alpha\rho}^{-1} (X - \mu_{\iota\sum})) \]  

(7)

We can sort size order of K gaussian distributions according to \( \omega_{\iota\alpha} \), and suppose the first B distributions as the background models.
We select the mixture of several Gaussian distributions to simulate each pixel’s historical record \{X_1, X_2, \ldots, X_t\}, and suppose the quantity of Gaussian distributions which is used to describe each point’s color distribution is \(K\) in total, which is denoted as respectively:

\[
\eta(X_i, \mu_j, \Sigma_j) = 1, 2, \ldots, K
\]

where, the subscript \(t\) denotes time. Therefore, the probability of the present pixel \(X_t\) is:

\[
P(X_t) = \sum_{i=1}^{K} \omega_i \eta(X_t, \mu_i, \Sigma_i)
\]

where, \(\omega_i\) is the weighted value of each single model, which indicates the reliable degree of using the present model to express pixel; \(\mu_i\) is the mean value of the single model, which represents the center of each single-peak distribution; \(\Sigma_i\) denotes the width value of the single-peak distribution of the single model, which expresses the instability degree of the pixel value; \(K\) expresses the quantity of the peak of the pixel value’s multi-distribution and the value of \(K\) depends on the pixel value’s distribution and meanwhile depends on the system’s computing power, which is set beforehand within 3-5. In this paper, we build the model dynamically, according to the distributions of each pixel value. We can find that, the above method is to express one pixel value’s observation with several single models. In order to make the model constantly get closer to the present pixel value’s distribution law, we need to update the model’s parameter for each newly obtained pixel value. The modified steps of its parameter are as the following:

1) As for each new pixel, we should detect whether it matches with the model at first and the detection method is:

\[
\begin{align*}
\text{matched} & \quad |X_t - \mu_j| < \lambda \sigma_j, \quad i=1, 2, \ldots, K \\
\text{unmatched} & \quad |X_t - \mu_j| \geq \lambda \sigma_j
\end{align*}
\]

where, \(\lambda\) is the constant set according to experiences. When we extract the samples from the overall quantity in normal distribution and about 95% of the samples are fallen in the range \((\lambda - 2 \sigma , \lambda + 2 \sigma)\), therefore, we usually set the value of \(\lambda\) within 2-3.

2) According to two different situations get from the first step, we can adopt different modified methods:

If a certain Gaussian (the \(k\) order number) of the multimode collection is matched with pixel \(X_t\), we need to update the Gaussian distribution’s weighted value, and the updated method is shown as the following:

\[
\omega_{k,j} = (1-\alpha)\omega_{k,j-1} + \alpha
\]

where, \(\alpha\) is the other constant to represent the background’s updating speed—the weighted value update rate, which is given with \([0, 1]\), and is set according to experience and the concrete condition. In order to reduce background noise, we set a relative small value of \(\alpha\), for example, 0.05.
The reason why we use the above formula to modify the parameter is that we hope the defined model can always truly simulate the distribution situation of the background pixel value in the latest moment. According to the model’s definition, the weighted value represents the occurrence probability of the nearest pixel value. Then when the newly obtained pixel value is matched with certain or several single models in this distribution, it means that the single model relatively meets the present pixel value’s distribution and therefore we need to increase its weighted value properly. The weighted value’s update rate \( \alpha \) indicates the weighted value’s modified quantity and the large \( \alpha \) realizes the fast modification.

When a certain single model is matched with a newly obtained pixel value, we need to modify its model parameter \( \mu_{i,j} \) and \( \sigma_{i,j} \). That is because when the newly obtained pixel value is matched with a single model, according to the probability distribution, it must influence the originally estimated probability distribution.

The modified equation is in the following:

\[
\sigma_{i,j}(x, y) = (1-\alpha)\sigma_{i,j}(x, y) + \alpha \sigma_i(x, y) \quad (17)
\]

\[
\mu_{i,j}(x, y) = (1-\alpha)\mu_{i,j}(x, y) + \alpha \mu_i(x, y) \quad (18)
\]

where, \( I_{i,j}(x, y) \) is the gray scale of the newly obtained video frame in the point \( (x, y) \).

If the new pixel value \( X_i \) is not matched with some one Gaussian distribution, we can consider that the new pixel value doesn’t make any contribution to the single model’s distribution and then we don’t need to change the Gaussian distribution’s parameter and only change its weighted value according to the following formula:

\[
\omega_{i,j} = (1-\alpha)\omega_{i,j-1} \quad (19)
\]

This indicates that only the Gaussian distribution’s weighted value which is matched with \( X_i \) can get increased, and other distributions’ weighted values are all decreased.

When not any a Gaussian distribution in the multimode collection is matched with the new pixel value \( X_i \), it means that the new distribution is generated and the distribution should be in the multimode collection. So we need to add a new single model and in the mean time remove a Gaussian distribution from the original model collection. The concrete method is to remove the Gaussian distribution with minimum weight in the present multimode collection and introduce a new Gaussian distribution in the multimode collection according to \( X_i \) and also set a relative small weighted value and relative large variance.

In the newly introduced Gaussian distribution according to \( X_i \), its weighted value is the minimum weight in the present multimode collection, and the mean value is the new pixel value and the variance is a given relatively large constant.

3) After the above modification, we need to process the weighted values of each single model in the models with the normalization method. As for the updating of the above weighted values, when a model is matched with a new pixel, because

\[
\sum_{i=1}^{k} \omega_{i,j-1} = 1 \quad (20)
\]

where,

\[
\sum_{i=1}^{k} \omega_{i,j} = (1-\alpha)\sum_{i=1}^{K} \omega_{i,j-1} + \alpha = 1 \quad (21)
\]

So we don’t need to carry on the normalization processing.

If a new model is generated, we will conduct the normalization processing of the weighted values of each original model:

\[
\omega_{k,j} = \frac{\omega_{k,j}}{\sum_{i=1}^{k} \omega_{i,j}} \quad (22)
\]

We can simulate the pixel value by the above models, that is, we should judge whether each new pixel value is the object pixel or background pixel or not, which means we are judging which single models are used to express the background pixel. The model to express background pixel in a mixture model should have these features: relative large weighted value and relative small variance.

We must consider these two aspects of factors. But it is difficult to say which parameter is more important than the other parameter to judge whether the single model is background model or not. We don’t need to discuss these two parameters’ influence on judging whether it is background model independently and only need to consider these two parameters’ relative size in the model collection in order to get the solution of the models belonging to the background. We can take the size of the relative value \( \omega/\sigma \) as the priority level of each Gaussian distribution.

The method to get the background pixel model is shown as the following:

1. Compute the priority level \( \omega/\sigma \) of each gaussian model.
2. Sort orders of each gaussian distribution from high to low in turn according to the size of the priority level \( \omega/\sigma \).
3. Select the first \( B \) gaussian distributions from \( K \) gaussian distributions as the background model and \( B \) is defined as the following:

\[
B = \arg \min_b \left( \sum_{i=1}^{b} \omega_i > T \right) \quad (23)
\]

where, \( T \) is considered as the minimizing measure of estimating background. By setting \( T \), we can select the best distribution of describing the background. If \( T \)’s
value is relatively small, the background model is usually single modal. And if this corresponds with the real situation, we can get satisfied effect only with the distribution which has the highest priority level. If T’s value is relatively large, the background model can describe the multimode situation caused by the background’s repeated changes, for example, the leaf’s floating and the water traces’ wave motion, etc. This will produce obvious effect on estimating the same pixel with two or more different colors in background. The background model built according to the first B gaussian distributions matches \( X_i \), with each of B gaussian distributions separately in the priority level order. If there is no any gaussian distribution which expresses the background distribution is matched with \( X_i \), this point is judged as the foreground point, otherwise as the background point, and then we can complete the object detection under the adaptive multi-gaussian model.

B. Illumination Variation Detection

Because GMM supposes that each pixel is independent separately, when the external illumination variation causes the scene change, it is quite possible to bring about false objects in large area which leads to misjudge. Through the observation of the illumination variation in the real video application, this paper will divide the illumination variation into two kinds: Sudden Variation and Slow Variation. By the analysis on the two variations, we introduce the illumination variation factor \( \theta_i \) to eliminate the illumination variation’s influence on the moving object’s detection:

\[
\theta_i = 1 - \frac{E_{r,i}}{E_i},
\]

\[
E_i = \sqrt{E_r^2 + E_b^2 + E_g^2}
\]

where, \( E_i \) expresses the present frame’s information entropy and \( E_r, E_b, E_g \) represents the present frame’s information entropy of each component separately. The image’s color is closely connected with the objects of them and different illumination pixel value’s distribution can reflect the environment’s illumination variation degree. Therefore, this paper adopts the color histogram to extract color feature. When the two detected images’ sample feature values are not the same but the difference of the two values is smaller than a certain threshold value, this indicates that their similarity is relatively high and the two images’ statistical distributions are similar. According to this principle, we propose the illumination variation detection method that is histogram matching algorithm in order to distinguish the illumination’s slow variation from sudden variation. Its formula is shown as the following:

\[
D(t,t-1) = 1 - \sum_{m=1}^{M} \min(H_i(m), H_{i-1}(m))
\]  

where, \( H_i \) represents the image’s histogram in t time and we process the normalization according to Equation(10), and get:

\[
D(t,t-1) = 1 - \frac{\sum_{m=1}^{M} \min(H_i(m), H_{i-1}(m))}{\sum_{m=1}^{M} \min(H_i(m))}
\]  

This paper uses \( D(t,t-1) \) to distinguish the illumination’s sudden variation and slow variation. The discrimination method is in the following equation:

\[
\begin{align*}
\text{Slow Variation, if } & D(t,t-1) \leq T \\
\text{Sudden Variation, otherwise}
\end{align*}
\]

where, \( T \) is the similarity matching threshold value. After distinguishing the illumination variation, we will update adaptively the learn rate \( \alpha \) according to the following Equation:

\[
\begin{align*}
\alpha = \alpha + \theta & \text{ if Slow Variation} \\
\alpha = 2\alpha & \text{ if Sudden Variation}
\end{align*}
\]

C. The Adaptive Gaussian Mixture Model

In traditional GMM, K’s value is a fixed constant, which builds K distributions for each pixel. But in real background, the quantity of each pixel’s modal distribution is not all equal. The relatively stable area is possible to be single modal, which can build model with one Gaussian component, and the relatively busy area need to build model with several gaussian components. As is shown in figure 1, it means the gray value statistical histogram of a certain video’s certain two pixel points in the time domain. From this, we find that the sample value’s gathering feature is possible to be single peak value or multi-peak value.

In order to reduce the redundant gaussian components and reduce the computing power, the literature [9] proposes an online iterative algorithm which introduces modal Dirichlet prior probability and according to the maximum a posteriori probability or the minimal information length standard result, automatically abandons or adds gaussian component’s quantity at the same time of estimating the parameter. This makes K’s value dynamically adapt each pixel’s multi-peak feature. The iteration formula of the modal weight is:

\[
\omega_{i+1} = (1 - \alpha)\omega_i + \alpha M_i - \alpha c
\]
where, \(c_T\) is a constant reflecting the dimension of the model parameter. To adaptively select the proper modal quantity is the object of model design, which can not only improve the stability of the model, but also save the computation time effectively and offer great help for detecting the system’s real time performance.

On the basis of the above algorithm in the literature [9] and through the adaptive adjustment on the recursive learn rate, this paper proposes an adaptive gaussian mixture model. The learn rate \(\alpha\) of the traditional gaussian mixture model is a constant, while the single learn rate cannot adapt dynamically the scene variation. This paper combines the illumination variation factor \(\theta_i\) to adjust the learn rate \(\alpha\) in real time according to formula(13) and meanwhile updates the learn rate \(\rho\) in formula (6) and (7) automatically.

V. THE EXPERIMENT RESULT AND ANALYSIS

In order to verify the proposed algorithm’s validity, this paper carries on the comparison experiment on several video sequences in different scenes. The experiment environment is in the PC of 2.5GHZ. This paper firstly makes comparison experiment aiming at the illumination variance’s influence on the detection system and then performs test on the algorithm’s running and processing time.

This paper selects the image sequence which can separately characterize illumination’s sudden variation and slow variation. Among them, GroudTruth is the true value of the test frame’s corresponding moving object binary image in each video database to verify the illumination variation adaptability of the algorithm proposed in this paper, as is shown in figure 2. The image sequence in figure 2(a) is LightSwitch; the image sequence in figure 2(b) is Time of Day; the image sequence in figure 2(c) is Highway.

In LightSwitch, the testing personnel make the experiment of switching the light on and off indoor to simulate the illumination’s sudden variation; In Time of Day, the testing personnel simulate the process of the sky’s gradient colors; In Highway, the swinging leaf and the road’s reflection to the illumination etc, lead to the illumination’s local sudden variation. From figure 2, we can find that, GMM’s adaptability of illumination variation is relatively low, which produces a large false foreground. The illumination variation’s adaptability of the algorithm in this paper and the algorithm from literature [9] (that is Z_GMM), have been improved in some way. Led by the adaptive learn rate \(\alpha\), the algorithm in this paper, is better than Z_GMM algorithm in the foreground’s segmentation result when in illumination sudden variation; while in illumination slow variation, the foreground’s segmentation result is similar to that of Z_GMM algorithm.

Since the Z_GMM algorithm can automatically select and describe the quantity of gaussian component of each pixel point, its processing time is similar to that of the algorithm in this paper. As in table 1, this paper compares the algorithm here with GMM through the image sequence of three different scenes’ images and the mean processing time of each frame between the algorithm in

![Figure 2](attachment:image.jpg)
this paper and the GMM algorithm. Among them, the fixed gaussian component quantity of GMM is 4, while the maximal gaussian component in this paper is 4, too. From table 1, we can find that, in the congested and complicated scene, the mean processing time of GMM algorithm is a little longer than the algorithm in this paper ; While in the simple scene, the mean processing time in this paper has been improved in some way.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>THE COMPARISON OF THE ALGORITHM’S PROCESSING TIME OF EACH FRAME (UNIT: MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM</td>
<td>17.86 17.67 17.06</td>
</tr>
<tr>
<td>The proposed method</td>
<td>16.09 9.56 6.28</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

The proposed method improves the algorithm of GMM from two major aspects. Firstly, for solving the influence on detection system brought by illumination variation, it can differentiate sudden illumination variation and slow variation with color histogram matching algorithm and update learning rate $\alpha$ adaptively by introducing illumination variation factors. Then, the convergence speed is improved by introducing model parameter updating counter $c_i$. It also can update learning rate $\rho$ when the model is updating, and the convergence speed will be improved. Combining with reference, the fixed number of Gaussian component can be adjusted automatically, in order to improve the real time performance of detection algorithm. And the robustness and stability of the detection effect have been confirmed in experiment.

REFERENCES