A Novel Object Detection Method Using Gaussian Mixture Codebook Model of RGB-D Information

Lujiang LIU¹, Gaopeng ZHAO*,¹, Yuming BO¹

¹School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China

Abstract — RGB-D sensors are being widely used in the computer vision world. Depth information is particularly attractive and suitable for object detection application as it is complementary with RGB information in solving many classic issues such as shadow interference, illumination change impacts and noise. In this paper a novel object detection method is proposed based on Gaussian mixture codebook model. Firstly, we establish the Gaussian mixture codebook model by combining the Gaussian model with codebook model by using the color data and the depth data. A four-dimensional Gaussian model based on R, G, B and D components is built in codewords. Then the object detection result is obtained based on the proposed model and the model updating method given. We then evaluated the method using publicly available datasets. The qualitative and quantitative experimental results show that the proposed method is more effective than the compared methods in complex scenes.

Keywords - object detection; gaussian mixture codebook model; RGB-D information

I. INTRODUCTION

In recent years, the accurate moving object detection in video sequences plays an important role in the applications of computer vision such as video surveillance, human computer interaction and so on [1]. The widely used approach of moving object detection is based on RGB color information. Researchers have proposed many effective detection algorithms, such as Bayesian decision rules [2], Mixture of Gaussians [3-4], and Kernel density estimation [5]. However, the color-based algorithms face many challenging problems including the following: vulnerable to illumination changes; shadows cast by moving objects: camouflage i.e., similar color between moving objects and the background [6] et al.

With the rapid development of depth data acquisition technology, moving object detection based on depth information is able to compensate for drawbacks in RGB data. But the depth data are usually noisy and unreliable. There are invalid results in the detection algorithm based on depth data alone [7]. On the base of the above considerations, the optimal algorithm, which fuses the RGB information and depth information, is attractive. So that the intrinsic limitations of single information can be counterbalanced and improved detection results can be obtained [8].

Recently, some researchers have proposed some methods based on RGB-D information to detect moving object. A logical operation “or” is used to combine the different foregrounds that respectively come from grayscale image and distance image [9]. The method can successfully cope with problems that the object has similar color with the background and its distance is close to the distance of the background, but it fails to overcome the edge noise. The authors extract foreground objects from depth image based on the method of region growing and use RGB information to refine the foreground object [10]. It not only solved the limitation of color camouflage, but also decreased the depth noise. However, it is less effective in complex scenes.

In this paper, we present a novel background subtraction algorithm, which aimed to solve the difficulty of adjusting the parameters and overcome the disadvantages only based on RGB information or depth information. In this algorithm, we combine the codebook model (CB) [11] with Gaussian model together. And a four-dimensional Gaussian model is built which based on R, G, B and D components in codewords. In this way, there is the characteristic of mixture of Gaussian model in codebook algorithm. The proposed method is more effective in complex scenes than the compared background subtraction algorithm. The results show a quantitative and qualitative improvement in the moving object detection application.

II. PROPOSED METHOD

Depth-based algorithm has strong robustness on sudden lighting changes, highlighted regions and shadows, which is difficult to the color-based algorithm. However, when the objects are closed to the background, depth-based codebook algorithm can classify the pixels to the background. An example is depicted in Figure 1, Figure 1 (a) is the original color image, Figure 1(b) is the corresponding depth image, Figure 1(c) is the result of codebook algorithm in [11], which only uses the color information; Figure 1(d) is the result of codebook algorithm in [11], which only uses the depth information. We can see that foreground objects are mistakenly detected in Figure 1(c) and Figure 1(d), due to there are shadows in Figure 1(a) and the close range between the book and the wall in Figure 1(b). Consequently, the detection result is not satisfied when the color or depth information is used individually.

In this paper, a Gaussian mixture codebook model, named as GMCB, is presented through combining the Gaussian model and codebook model based on RGB-D information. The proposed background subtraction method based on the GMCB includes background model.
construction and foreground detection. The background model construction method is illustrated in section II.A; and the foreground detection and model updating method is illustrated in section II.B.

The Gaussian mixture codebook model is the use of quantitative techniques in the long-term observation sequence to build background model. It builds a codebook model for each pixel. In the process of initialization algorithm training, \( X \), which is the pixel value of a single pixel in a training sequence, is consisted of \( N \) RGB-D vectors: \( X = \{ x_1, x_2, ..., x_N \} \). Let \( \zeta = \{ c_1, c_2, ..., c_L \} \) represent the codebook that is composed of \( L \) codewords. For each pixel, the number of codewords may be not the same depend on its sample variation. Each codeword \( c_i \ (i = 1...L) \) consists of a twelve-tuple as formula (1).

\[
 c_i = ( \mu_{R,i}, \mu_{G,i}, \mu_{B,i}, \mu_{D,i}, \\
 \sigma^2_{R,i}, \sigma^2_{G,i}, \sigma^2_{B,i}, \sigma^2_{D,i}, \\
 f_i, \lambda_i, p_i, q_i )
\]

In equation (1), every symbol is expressed as follows:
- \( \mu_{R,i}, \mu_{G,i}, \mu_{B,i}, \mu_{D,i} \) --- mean value of R, G, B, D for each pixel;
- \( \sigma^2_{R,i}, \sigma^2_{G,i}, \sigma^2_{B,i}, \sigma^2_{D,i} \) --- mean square deviation of R, G, B, D for each pixel;
- \( f_i \) --- the frequency with which codeword \( c_i \) has occurred;
- \( \lambda_i \) --- the maximum negative run-length, defined as the longest interval during the training period that the codeword has not recurred;
- \( p_i, q_i \) --- the first and last access times, respectively, that the codeword \( c_i \) has occurred;

The condition, which an incoming pixel \( x = (R_i, G_i, B_i, D_i) \) is matched successfully with the codeword \( c_i \), is defined as formula (2).

\[
 b(x, c_i) = 1 \quad (2)
\]

We define 1 for matched correctly and 0 for matched incorrectly. The matching condition is defined as formula (3).

\[
 b(x, c_i) = \begin{cases} 1, & (Z - \mu^{z})^2 \leq \alpha \sigma_{z,m}^2 \\ 0, & otherwise \end{cases} \quad (3)
\]

Where \( Z \in \{ R, G, B, D \} \), \( z \in \{ R, G, B, D \} \).

According to the statistics, the probability of a random variable, which obeys Gaussian distribution and is in \((\mu - 2.58\sigma, \mu + 2.58\sigma)\), is 99.7%. Therefore, prior parameter \( \alpha \) is 2.58^2.

Mean and variance update methods are defined as the formula (4) and formula (5).

---

**Figure 1.** Detection results by codebook algorithm [11]

**A. Background Modeling**

Based on the assumption that the pixel value of the same position in the video sequence can be modeled to Gaussian distribution, we propose a novel Gaussian mixture codebook model using Gaussian model to improve codebook model based on RGB-D information. The whole codebook has the characteristics of the mixture of Gaussian as the Gaussian model is built in the R, G, B, D channel separately.
The detailed GMCB background model construction method is shown in Table I.

<table>
<thead>
<tr>
<th>Steps</th>
<th>GMCB background model construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$L \leftarrow 0$, $\zeta \leftarrow \Phi$ (empty set)</td>
</tr>
<tr>
<td></td>
<td>for $t=1\ldots N$ do</td>
</tr>
<tr>
<td></td>
<td>(1) In $\zeta = {c_1, c_2, \ldots, c_L}$, finding the codeword $c_i$ that match to $x_t$ satisfy the following conditions: $b(x_t, c_i) = 1$.</td>
</tr>
<tr>
<td></td>
<td>(2) If $\zeta = \Phi$ or there is no matched codeword, then $L \leftarrow L + 1$, a new codeword $c_L$ is created and added to $\zeta$.</td>
</tr>
<tr>
<td></td>
<td>(3) If not, update the matched codeword $c_i$, $c_i \leftarrow {R_i, G_i, B_i, D_i, 0, 0, 0, 0, 1, t-1, t, t}$.</td>
</tr>
<tr>
<td>2</td>
<td>$c_i \leftarrow \left{ \frac{f_i \mu_{R_i} + R_i}{f_i + 1}, \frac{f_i \mu_{G_i} + G_i}{f_i + 1}, \frac{f_i \mu_{B_i} + B_i}{f_i + 1}, \frac{f_i \mu_{D_i} + D_i}{f_i + 1}, \frac{f_i \sigma_{R_i}^2 + (R_i - \mu_{R_i})^2}{f_i + 1}, \frac{f_i \sigma_{G_i}^2 + (G_i - \mu_{G_i})^2}{f_i + 1}, \frac{f_i \sigma_{B_i}^2 + (B_i - \mu_{B_i})^2}{f_i + 1}, \frac{f_i \sigma_{D_i}^2 + (D_i - \mu_{D_i})^2}{f_i + 1}, f_i + 1, \max {\lambda_i, t - q_i}, p_i, t}$.</td>
</tr>
<tr>
<td>3</td>
<td>For each codeword $c_i (i = 1, \ldots, L)$, update $\lambda_i$ by $\lambda_i \leftarrow \max {\lambda_i, (N - q_i + p_i, t - 1)}$.</td>
</tr>
<tr>
<td>4</td>
<td>For each codeword $c_i (i = 1, \ldots, L)$, Delete the $c_i$ which $\lambda_i &gt; N/2$. $L \leftarrow L - 1$.</td>
</tr>
</tbody>
</table>
| 5     | Finally, the final model is $C = \{c_i | 1 \leq i \leq L \land \lambda_i \leq N/2 \}$.

B. Foreground Detection
The foreground detection can be depicted as a classification question and get the detection result. An incoming pixel is classified into foreground or background according to the formula (2) and (3) in section II.A. The detection process is given in formula (6).

$$BGS(x) = \begin{cases} FG & \text{if } \zeta = \Phi \text{ or there is no matched codeword}, \\ BG & \text{otherwise} \end{cases}$$

Where $FG$ represents the foreground, $BG$ represents the background, $BGS(x)$ represents the detection result.

C. Model Updating
When the pixel is classified to foreground, the model is not need to update. But if the pixel is classified to background, the model updating is executed to resolve the effect of the background change. The model updating process is given in formula (7). The $x_t$ is used to update the matched codeword $c_i$. The formula (7) is the same as the step 2 (3) in Table I.

$$c_i \leftarrow \left\{ \frac{f_i \mu_{R_i} + R_i}{f_i + 1}, \frac{f_i \mu_{G_i} + G_i}{f_i + 1}, \frac{f_i \mu_{B_i} + B_i}{f_i + 1}, \frac{f_i \mu_{D_i} + D_i}{f_i + 1}, \frac{f_i \sigma_{R_i}^2 + (R_i - \mu_{R_i})^2}{f_i + 1}, \frac{f_i \sigma_{G_i}^2 + (G_i - \mu_{G_i})^2}{f_i + 1}, \frac{f_i \sigma_{B_i}^2 + (B_i - \mu_{B_i})^2}{f_i + 1}, \frac{f_i \sigma_{D_i}^2 + (D_i - \mu_{D_i})^2}{f_i + 1}, f_i + 1, \max \{\lambda_i, t - q_i\}, p_i, t\}.$$
is adopted to assist image processing. We provide the visual comparison of the above methods and the quantitative results are given by using the hand-segmented ground truth. The first 50 frames in each sequence are used to background model construction and the other frames are used to detection evaluation.

B. Qualitative Analysis

1) Scene 1: Target and Background Similar Distance

In the scene 1, a person hands a book keeps away from the wall. We select the 78th frame, the 135th frame and the 142th frame for visual comparison as shown in Figure 2 from column 1 to column 3, among them the 78th frame is nearest from the background and the 142th is furthest from the background.

As can be seen in Figure 2, there are large detection error in the result of the CBColor method because of the shadow of the book. The CBDepth method can solve this shadow problem, but object cannot be detected in the 78th frame because of the similar distance between object and background. The CB4D method shows better effect by using the color and depth information. But there is still some mistakes. Although the result of the GMCB method in the 78th frame is not satisfied, the GMCB method gets the better result than the others.
LUJIANG LIU et al: A NOVEL OBJECT DETECTION METHOD USING GAUSSIAN MIXTURE CODEBOOK …

2) Scene 2: Target and Background Similar Color

As shown in Figure 3, a man is holding a white box and two blue books and goes through the scenes. The color of box is similar to the wall and the color of books is similar to the trash. The 290th frame is selected for visual comparison as shown in Figure 3(a) and Figure 3(b). Figure 3(d)-(g) are the detection results of the 290th frame with the different methods. Compared with the ground truth in Figure 3(c), we can find there are many artifacts in the result of the CBColor method in Figure 3(d). The detection result is better in Figure 3(e), which is the result of the CBDepth method, but the white box and books are only partially detected. The GMCB method detects the white box effectively in Figure 3(g). The effect of the proposed GMCB method is obviously effective and it is robust to the difficult situations.

3) Scene 3: Sudden Illumination Changes

There are sudden illumination changes in scene 3. The 368th frame is selected for visual comparison as shown in Figure 4. We can find that CBColor method and CBDepth method cannot adapt to sudden changes in illumination. There are a lot of noises in CB4D method. The GMCB method gets the better detection result than the others.
C. Qualitative Analysis

Precision ($P$), recall ($R$) and $F$-measure ($F$) are the common evaluation criteria of object detection. Recall is the true positive; precision is the ratio between the number of correctly detected pixels and the total number of pixels marked as foreground; $F$-measure is a successful combination of $P$ and $R$ to comprehensively evaluate the performance of the algorithm.

The values of $P$, $R$, $F$ can be computed as $P = \frac{TP}{TP + FP}$, $R = \frac{TP}{TP + FN}$, $F = \frac{2PR}{P + R}$. The $TP$ (True Positive) value is the number of pixels as the moving object is correctly detected. $FP$ (False Positive) is misidentified as the background pixels. $FN$ (False Negative) is the number of pixels to be mistaken for the moving object. This $F$-measure not only offers a trade-off between the ability of an algorithm to detect foreground and background pixels, but also provides a general evaluation of robustness of the algorithm. In general, the value of $F$ is higher, the better the performance.

<table>
<thead>
<tr>
<th>Method</th>
<th>$F$ value</th>
<th>Scene 1</th>
<th>Scene 2</th>
<th>Scene 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBColor</td>
<td>0.6002</td>
<td>0.6438</td>
<td>0.4222</td>
<td></td>
</tr>
<tr>
<td>CBload</td>
<td>0.7863</td>
<td>0.7982</td>
<td>0.7331</td>
<td></td>
</tr>
<tr>
<td>CB4D</td>
<td>0.7705</td>
<td>0.7965</td>
<td>0.6586</td>
<td></td>
</tr>
<tr>
<td>GMCB</td>
<td>0.8102</td>
<td>0.8229</td>
<td>0.8127</td>
<td></td>
</tr>
</tbody>
</table>

For each scene, by computing the $P$ value, $R$ value, and $F$ value for every detection results obtained by above method, the final $F$ values are given by averaging the corresponding values. The results are shown in Table II.

It can be observed in Table II that the detection performance based solely on RGB or depth information is poorer and even appear failure. The detection performance of CB4D is also the case. However, the proposed GMCB method can keep a high detection performance in all the test complex situations.

IV. CONCLUSIONS

The problem of object detection is a well-known problem, but still far from being solved. In this work we proposed the Gaussian mixture codebook model of RGB-D information. Depth information is as the fourth channel in codeword, and the Gaussian model is built in the RGBD channels separately. We evaluated the method on the publicly dataset, which includes the complex scenes such as similar distance, similar color and sudden illumination changes. Experimental results show that the proposed method obtained a satisfying detection results on accuracy and robustness.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61203266).

REFERENCES


