MULTIPLE SAMPLE GROUP PAIRS’ GRAPH EMBEDDING FOR TRACKING

Lin Ma¹, Weiming Hu¹, Xiaoqin Zhang²
¹National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China
²College of Mathematics & Information Science, Wenzhou University, China
E-mail: lin.ma@ia.ac.cn

ABSTRACT

Abstract: This paper presents a new method which uses graph embedding and foreground-background patch pairs to perform object tracking. We first use particle filter to sample some particles. Then we evaluate each particle based on graph embedding and foreground-background patch pairs. For each particle, we use a two-layer model to represent the object, i.e. the inner layer (object layer) and the outer layer (background layer). Both the two layers are divided into patches. We cluster the foreground patches to several classes. Each class forms one sample group pair with the background patches. We perform graph embedding on multiple sample group pairs to discriminate the foreground and the background. Experimental results showed that our method tracked the objects efficiently.

Index Terms— Graph embedding, tracking, histogram

1. INTRODUCTION


Graph embedding (GE) was an important learning method which was able to unify different dimensionality reduction methods [7, 8]. Tiwari [9] presented a unified data fusion framework, Semi Supervised Multi Kernel Graph Embedding (SeSMiKGE), and used the new method to identify aggressive prostate cancer via Magnetic Resonance Imaging and Spectroscopy. In object tracking, generally the background objects had some discrimination with the foreground objects. The discrimination was able to be used to obtain the state of the tracked object [10, 11]. Reddy [12] segmented and tracked the foreground object based on patch-based background modeling.

This paper adopts the framework of particle filters to perform tracking. We first sample some particles, and then use the information of GE and the foreground-background patch pair (FB) to evaluate each particle. We divide the foreground and background to patches and cluster the foreground patches to several classes. Each class forms a sample group pair with the background patches. We perform GE on multiple sample group pairs to discriminate the foreground and the background. In computing the projection directions, samples far from the discriminative plane have more importance than samples near the discriminative plane. We propose a new method to revise the samples’ weights according to the distances between the samples and the discriminative plane. The flowchart of our method is shown in Fig. 1. The key novelties of our paper are as follows:

1) We perform clustering on foreground patches before GE. This is useful to present the difference between foreground patches, and is able to make the tracking more accurate.
2) We propose a new way to define the samples’ weights in computing the projection directions according to the distances between the samples and the discriminative plane.
3) We combine GE and FB together to perform tracking and are able to track the object more effectively.

The rest of our paper is organized as follows: Section 2 shows the particle filter we use in this paper. In Section 3, we evaluate the particles with the new GE method. In Section 4, we use the FB model to evaluate the particles. Experimental results are shown in Section 5, and conclusion and future work are shown in Section 6.

Figure 1: Flowchart of our method

2. PARTICLE FILTER

The particle filter is formed based on Bayesian formula. Given the observation sequence \( O_{1:t+1} \), the posterior probability density of the object state can be defined as
where \( (tx, ty, sx, sy) \) represents the object state at time \( t \).

We warp each particle to a standard 32×32 sub-image. We define \( (tx, ty) \) as the coordinate of the bottom left point of the state rectangle in the frame image, and define \( sx, sy \) as the ratio of the object’s width and height to the standard sub-image’s width and height respectively.

We assume that each component, e.g. \( tx \), of the particle \( X_{i+1} \), is independent of the others, and each component conforms to the Gaussian distribution, i.e. \( X_{t_i}(i) \sim N(X(i); X(i,t); \sigma^2_{tx}), t = 0, \ldots, 3 \). Here \( X_i(i) \) represents the \( i \)th component, e.g. \( tx \), of the state \( Xt, \sigma^2_{tx} \) is a constant. Detailed presentation of particle filter can be found in [3]. We select the particle with the largest likelihood \( \pi_t = p(O_t | X_t) \) as the optimal particle.

3. GRAPH EMBEDDING

3.1. Graph embedding

GE is a framework of dimensionality reduction. PCA, LDA, etc. can all be combined in this framework [7, 8]. Let \( x_t \in \mathbb{R}^d, i = 0, \ldots, N – 1 \) be \( \mathbb{R} \)-dimensional samples (here \( d=64 \)), and \( y_t \in \{0, \ldots, C-1\} \) is \( x_t \)’s class label. \( n_c \) is the number of samples belonging to class \( c \), and satisfies \( \sum_{c=0}^{C-1} n_c = N \). We construct an undirected graph \( G = \{ \mathcal{X}, W \} \) with the vertex set \( \mathcal{X} \), and the similarity matrix \( W \), where \( \mathcal{X} = \{x_0, \ldots, x_{N-1}\} \) is the sample set (each patch is a sample in this paper). The element \( w_{ij} \) of \( W \) represents the similarity between vertex \( x_i \) and \( x_j \). We define the element \( d_i \) of the diagonal matrix \( D \) as

\[
d_i = \sum_{j \neq i} w_{ij}^{2}
\]

Then the Laplacian matrix \( L \) is defined as

\[
L = D - W
\]

The intention of GE is to obtain the optimal low dimension representation of the sample, given the similarity matrix and class labels. The projection vectors is obtained by solving

\[
P^* = \arg \min_{P \in \mathcal{P}} \sum_{i \neq j} |z_i - z_j|^2 \quad \text{subject to} \quad \text{tr}(P^T LP^*) = \text{tr}(P^T Z^T L Z)
\]

where \( z_t = P^T x_t, \text{tr}(v) \) represents the trace of the vector \( v \).

We define \( T_{i,t} \) as average of patch \( i \) of frame \( 0, \ldots, t \). To make use of the similarities between foreground patches, we cluster the foreground patches \( T_{i,t}, i = 0, \ldots, 15 \) to 2 sample groups with k-means clustering during the training process. We choose the two foreground patches with the largest distance from each other as the two beginning samples. The two sample groups are updated adaptively. We define \( \pi^{(k)}_j \) as average of patch \( j \) of recent 5 frames, and then define

\[
\alpha_k = \left(1/\sqrt{2\pi V(k)}\right) \exp\left(-\frac{|i_1 - i_2|^2}{2V(k)}\right)
\]

and

\[
\beta_i = \left(1/\sqrt{2\pi V(k)}\right) \exp\left(-\frac{|i_1 - i_2|^2}{2V(k)}\right)
\]

Then the entry of \( W_k \) is defined as

386
is defined in this way because generally foreground patches are more steady than background patches.

3.3. Weights of samples

In computing the projection vectors according to Eq. (6), normally the samples with different distances from the discriminative plane have different importance. We define

\[ \omega_{ij}^{(k)} = \alpha_{k, i} \beta_{j} \tag{10} \]

where \( \alpha_{k, i} \) is the \( i \)-th column vector of \( \mathbf{X}^{(k)} \). We update \( \omega_{ij}^{(k)} \) with \( \omega_{ij}^{(k)} = [l_{k, i} a_{x_{k,0}}, \ldots, l_{k, i} N_{k}^{(k)} + 1] \) to change the samples’ projection values, where \( l_{k, i}, i = 0, \ldots, N_{k}^{(k)} + 1 \) is coefficient. We assume that the new projection value of \( x_{k, i}^{*} \) is \( g(p_k^T x_{k, i}^*) \), where \( g(x) \) is a function. At the beginning, we make all the samples have the same coefficient 1. Then we compute the coefficient \( l_{k, i} \) of \( x_{k, i}^{*} \) iteratively. As \( l_{k, i} p_k^T x_{k, i} - g(p_k^T x_{k, i}^*) \), each time when the new \( p_k, p_i \) are obtained, we set

\[ l_{k, i} = g(p_k^T x_{k, i}^*) / g(p_k^T x_{k, i}) \tag{11} \]

In this paper, we define \( g(x) = x^E \), where \( E \) is a constant. Reducing \( E(>0) \) improves the importance of the samples near the discriminative plane, while enlarging \( E(>0) \) increases the importance of samples far from the discriminative plane. The new multiple sample group pairs’ GE method is summarized in Algorithm 1.

Algorithm 1. The proposed new GE method

Input: \( x_{i}, i = 0, \ldots, N - 1; y_{i}, i = 0, \ldots, N - 1 \)

Steps: 1. For the initial 5 frames, cluster foreground patches to 2 groups, and form 2 sample group pairs. For other frames, redistribute the foreground patches.

2. Compute \( W_k \) according to (10).

3. Compute \( p_{k}, k = 0, 1; \)

   1) Initialize the projection weight \( l_{k, i} = 1 \)
   2) Obtain \( p_k \) according to (6)
   3) Revise \( l_{k, i} \) according to (11), and do 2) of Step 3 again, until reach the maximum iteration time (2 in this paper).

Output: \( p_0, p_1 \)

4. FOREGROUND-BACKGROUND PATCH PAIRS

Normally the object boundary patches have large contrasts in appearance to their neighbour background patches, which form the FB (Fig. 2(b)). The contrast can be used to evaluate the particle.

There are 16 FBs (with shared edges). For patch pair \( i \), the background patch histogram \( h_{b, i} \) and its corresponding foreground patch histogram \( h_{f, i} \) usually have large contrast (Fig. 2(c)(d)). We define \( N_0 \) as the bin number (\( N_0 = 16 \) in this paper), and define the coefficient

\[ \rho(h_{f, i}, h_{b, i}) = \sum_{k=1}^{N_0} \sqrt{h_{f, i}(k) h_{b, i}(k)} \tag{12} \]

Figure 2: FBs. (a) Grid. Red rectangle: state rectangle. Pink rectangle: contain background and foreground. Yellow line: separate patches. White rectangle: an FB. (b) The two patch coordinates to the white rectangle in (a). Left: background patch. Right: foreground patch. (c) Background patch histogram of (b) left. (d) Foreground patch histogram of (b) right.

The likelihood corresponding to FB (i.e. the combined evaluation of the 16 histogram pairs) is defined as

\[ p(O_i, f_k | X_t) \propto \exp \left( \frac{1}{16} \sum_{i=1}^{16} (-\rho(h_{f, i}, h_{b, i})) \right) \tag{13} \]

The combined likelihood of \( X_t \) is defined as

\[ p(O_i, X_t) \propto p(O_i, f_k | X_t) p(O_i, f_k | X_t) \tag{14} \]

5. EXPERIMENTS

5.1 Experiments Setting

We tested our method on 3 videos. The experiments were done using VC++6.0 on a computer with CPU at 2.53 GHz, on the Win XP OS. For each experiment of our method, we updated the system every 5 frames, and used 150 particles per frame during tracking. The running time of our method was around 0.035 sec/frame. We used the sum of distances between the 4 state rectangle edges and the ground truth to represent the tracking error. The test videos were from EC Funded CAVIAR project/IST 200137540 [13] and [14].

5.2 Performances

In Fig. 3, we tracked a walking woman and the head of a girl. We defined that \( E=1/3 \) represented \( g(x) = x^{1/3} \), \( E=1 \) represented \( g(x) = x \), \( W=0 \) represented only using GE, \( W=1 \) represented using both GE and FB. From Fig. 4, we saw that combining GE with FB was able to improve the tracking efficiently. Compared \( E=1/3 \) with \( E=1 \), we found that when only using GE, \( E=1 \) tracked the objects more accurate. But when combined with FB, \( E=1/3, W=1 \) tracked the object more accurate than \( E=1, W=1 \). That was because \( E=1/3 \) tended to reduce the importance of the patch samples far from the discrimination plane. However, when the tracking was not accurate, the confidences of the samples near the discrimination plane were reduced and thus the projection vector computed by \( E=1/3 \) was less accurate than \( E=1 \). But
FB was able to improve the tracking accuracy, and the confidences of patch samples near the discrimination plane were improved. Then $E=1/3$ was able to discriminate more accurately the samples near the discriminate planes and track the objects more accurately than $E=1$. The error maps of Fig. 3 were shown in Fig. 5 (a)(b), and the average errors of Fig. 3 were shown in Table 1. $E=1/3$, $W=1$ was adopted by GE+FB in Fig. 4.

In Fig. 4, we tracked a girl sitting in a moving car and compared our method with LDA and SMOG. By clustering the foreground patches into two groups, we were able to divide the foreground and background patches more accurately. From Fig. 4, we saw that the face patches and the hair patches were clustered into 2 groups with the first 5 frames of the image sequence. The object appearance changed largely due to the light variance. LDA and SMOG were influenced by the appearance variance and by the similarity between foreground patches and background patches. However, GE considered the confidences of foreground patches and background patches in inter-class similarity matrix, and by combining FB our method (GE+FB) was able to track the object more accurately than LDA and SMOG. The error map of Fig. 4 was shown in Fig. 5, and the average errors of GE+FB, LDA and SMOG were 18.7, 22.5, 25.3 and 24.9 respectively.

Table 1. Average errors of the 4 ways in Figure 3

<table>
<thead>
<tr>
<th>$E=1/3, W=0$</th>
<th>$E=1/3, W=1$</th>
<th>$E=1, W=0$</th>
<th>$E=1, W=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 3(a)</td>
<td>27.5</td>
<td>23.6</td>
<td>26.2</td>
</tr>
<tr>
<td>Fig. 3(b)</td>
<td>26.7</td>
<td>22.5</td>
<td>25.3</td>
</tr>
</tbody>
</table>

This paper has presented a new method which uses GE and FB to perform object tracking. We clustered the foreground patches to 2 groups and formed 2 sample group pairs. We revised the samples’ projection weights according to the samples’ projection distances from the discriminative plane. The experiments validated our method’s efficiency. In the future research, we will go on with finding more efficient ways to represent the contrast between the foreground and the background. In another part, we will go on with the research of constructing more efficient similarity matrix.

This work is partly supported by NSFC (Grant No. 60825204, 60935002, 61100147), Beijing Natural Science Foundation (4121003), and Zhejiang Natural Science Foundation (Y12F030055).

7. REFERENCES


