LOGISTIC DYNAMIC TEXTURE MODEL FOR HUMAN ACTIVITY AND GAIT RECOGNITION

Changhong Chen1,*, Jimin Liang2, Xiuchang Zhu1

1College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications
2School of Life Science and Technology, Xidian University
*chenchh@njupt.edu.cn

ABSTRACT

In this paper, a logistic dynamic texture model (LDT) is proposed to characterize binary image sequences. Dynamic texture model (DT) is one of the most efficient and successful methods in modeling dynamic sequences. It learns the parameters through a closed-form solution and commonly uses principal component analysis (PCA) to obtain the observation function. PCA assumes a Gaussian distribution over a set of observations. However, the binary image sequences subject to Bernoulli distribution. The LDT introduces logistic PCA to learn the observation function. The proposed model is capable of describing the binary image sequences accurately by processing the pixels of 1 and 0 separately. The model is demonstrated by image reconstructing and activity/gait recognition experiments. Experimental results illustrate the effectiveness of our model.

Index Terms—logistic dynamic texture model; logistic PCA; activity recognition; gait recognition

1. INTRODUCTION

Human motion analysis [1, 2] has been one of the most active research fields. Human activity and gait recognition are two important aspects of human motion analysis. In order to eliminate the impact of background and the color or texture of the clothes, the human body needs to be segmented and represented as a binary silhouette. There is a temporal coherence intrinsic in the images of the same sequence. Our aim is to building a model which can capture the dynamic nature from the binary activity or gait sequence. The recognition problem will be simplified to the comparison between models.

Dynamic texture model (DT) [3] is a generative stochastic model that treats the video sequence as a sample from the output of a linear dynamical system. This model is simple and efficient to capture the essence of moving scenes that exhibit certain stationary properties in time. It applies to many domains [3-5]. DT learns the parameters through a closed-form solution and commonly uses principal component analysis (PCA) to get the observation function. It is successful to model texture sequences, but its performance in modeling binary image sequences is disappointed. The major reason is that PCA assumes a Gaussian distribution over a set of observations, while the binary image sequences subject to Bernoulli distribution. In order to avoid imposing DT on binary image sequence directly, some work [6-8] extracted features from binary image sequence. Bissacco [6] introduced a representation with the projection features. Bissacco [7] computed the affine invariant moments on the binary image and normalized them to form the feature vector. Mazzaro [8] measured the angles of the shoulder, elbow, hip and knee joints. Feature extraction or representation loses some useful information and decreases the accuracy of the binary image sequence. Therefore, these methods can only used on activity recognition of small database or classifying gait styles.

In this paper, we propose a logistic dynamic texture model (LDT) to directly model binary data. This model introduces logistic PCA [9] to learn the observation function. Logistic PCA assumes a Bernoulli distribution over a set of observations and processes the pixels of 1 and 0 separately, when decomposing the binary image sequence. The property of logistic PCA makes LDT more reliable. We evaluate the efficiency of the proposed model through image reconstruction and apply it to activity and gait recognition. Besides DT, we also compared it with kernel dynamic texture model (KDT) [10]. Experimental results demonstrate the LDT is superior to DT and KDT for binary image sequence description.

2. LOGISTIC DYNAMIC TEXTURE MODEL

We first briefly review dynamic texture model (DT) and then present logistic dynamic texture model (LDT) in detail. In the third part, we reconstruct the binary image sequence.

2.1. Dynamic texture model

DT is first proposed by Doretto [3]. It has an underlying assumption that individual images are realizations of the output of a dynamical system driven by an independent and identically distributed (IID) process. The visual components and underlying dynamics are represented as two stochastic processes, as shown in Fig.1. The model is represented as:
where $A$ is the state-transition matrix and $C$ is the observation matrix. The state and observation noises submit to Gaussian distribution, which are given by $v_t \sim N(0, Q)$ and $w_t \sim N(0, r I_p)$, respectively.

![Fig.1 Topology structure of DT](image)

The choice of matrices $A, C, Q, r$ is not unique and a closed-form solution was also proposed in [3]. The closed-form solution commonly uses principal component analysis (PCA) to get the observation function. The columns of $C$ are the principal components of the image sequence and the state vector $x_t$ is a set of PCA coefficients.

### 2.2. Logistic dynamic texture model

Although the DT is efficient to capture natural texture sequences, it is unsuitable to model the binary image sequences. The binary image sequence subject to Bernoulli distribution, while PCA used in the close-form solution assumes a Gaussian distribution over a set of observations. To tackle the binary image sequence, we propose an extension of DT, namely LDT. This model assumes that the individual binary images are realizations of the output of the model. It can also be represented by equation (1) and Fig.1.

The parameters of LDT are trained as follows:

**Step 1:** Conducting logistic PCA [9] on binary image sequence $Y = [Y_1, Y_2, \ldots, Y_N]$. Suppose coefficient matrix is $U$ and basis vector matrix is $V$. The initial values of $U$ and $V$ are stochastically given. The bias vector is neglected in this procedure for simplicity. Matrices $U$ and $V$ are updated by the ALS separately. When updating $U$, we fix $V$ and obtain a simple update rule that stores the basis vectors of the log-odds matrix $\Theta$.

$$\Theta_{nd} = \sum_l U_{nl} V_{ld} + \Delta_d,$$  \hspace{1cm} (2)

An intermediate quantity $T$ is calculated as:

$$T_{nd} = \frac{\tanh(\Theta_{nd} / 2)}{\Theta_{nd}},$$  \hspace{1cm} (3)

Matrix $U_{nl}$ is obtained by solving the $L \times L$ set of linear equations:

$$\sum_l \left( \sum_d T_{nd} V_{ld} V_{ld} \right) U_{nl} = \sum_d (2Y_{nd} - 1) V_{ld}.$$  \hspace{1cm} (4)

Similar update is implemented for matrix $V$ with $U$ holding.

**Step 2:** Calculating the parameters of the LDT according to the decomposition values. The coefficient matrix $U$ and the basis vector matrix $V$ are regarded as the observation and state matrices of the model. The state-transition matrix is calculated from the state matrix.

$$\hat{C} = U, \quad \hat{x}_1 \cdots \hat{x}_N = V,$$

$$\hat{A} = \left[ \hat{x}_2 \cdots \hat{x}_N \left\| \hat{x}_1 \cdots \hat{x}_{N-1} \right. \right].$$  \hspace{1cm} (5)

$\hat{Y}_t$ is the estimated observation obtained by:

$$\hat{Y}_t = \hat{C} x_t.$$  \hspace{1cm} (6)

$r$ is calculated as:

$$r = \frac{1}{DN} \sum_{i=1}^N \left\| \hat{Y}_i - \hat{Y}_i \right\|^2.$$  \hspace{1cm} (7)

$Q$ is given by:

$$Q = \frac{1}{N-1} \sum_{i=1}^{N-1} \left( x_i - A x_{i-1} \right) \left( x_i - A x_{i-1} \right)^T.$$  \hspace{1cm} (8)

### 2.3. Binary image reconstruction

The images are reconstructed as follows:

**Step 1:** Calculating the state at time $t$. Suppose $M$ is an arbitrary matrix and $M \in R^L$. The state $x_t$ is:

$$x_t = A x_{t-1} + Q M.$$  \hspace{1cm} (9)

**Step 2:** Obtaining the image vector. Suppose $P$ is a arbitrary matrix and $P \in R^B$. The image vector $y_t$ is:

$$y_t = C x_t + r I_d P.$$  \hspace{1cm} (10)

In order to quantify the inaccuracy, we calculate the reconstruction error rates. The error rate $e_t$ of the $t^{th}$ frame is:

$$e_t = \frac{1}{N} \sum \left| Y_t - \hat{Y}_t \right|.$$  \hspace{1cm} (11)

where $Y_t$ is the original image, $\hat{Y}_t$ is the reconstruction image and $N$ is the point number of the image. Fig.2 shows the average reconstruction error rates. The error rates of LDT increase slower. The average error rate of dynamic texture is 4%, while that of logistic dynamic texture is 2.13%.
3. EXPERIMENT RESULTS

We conduct experiments on human activity and gait databases to evaluate the proposed LDT. The models of the training and test sets are compared by Martin distance [11].

3.1. Activity recognition

The activity database in [12] is employed. It contains 81 videos of 9 people. Every people perform 10 activities: bend, jack, jump, pjump, skip, side, run, walk, wave1 and wave2.

The leave-one-out method is adopted in the experiments. Each time the cycles of the same activity belongs to the same people are left out, while the others are used for training. Two cycles are obtained for each activity. 2*9 experiments are conducted for each activity. The cycle length of different activities varies from about 10 to 35 frames. Therefore, the dimension of latent space is set to a small value (9 in the experiments). The iteration times of the LDT are set to be 30.

Table I: Confusion matrix of DT, KDT and LDT on jump and pjump

<table>
<thead>
<tr>
<th></th>
<th>jump</th>
<th>pjump</th>
<th>bend</th>
<th>jack</th>
<th>skip</th>
<th>side</th>
<th>run</th>
<th>walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>jump</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>KDT</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LDT</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>pjump</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KDT</td>
<td>0</td>
<td>1K</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LDT</td>
<td>0</td>
<td>1K</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The recognition rates of bend, jack, skip, side, run, walk, wave1 and wave2 activities achieve 100%, whichever model is used. The mistakes occur on jump and pjump. We only show the results on jump and pjump activities Table I. Kernel dynamic texture model (KDT) has relative lower performance on jump. DT and LDT have similar performance and they wrongly take jump as run and walk. When pjump activity is tested, DT has 5.5% probability of taking it as jack. LDT performs better than DT and KDT.

3.2. Gait recognition

The CMU MoBo gait database [13] contains 25 individuals walking with two different speeds. The fronto-parallel sequences are adopted. Two experiments are set up as follows:

(a) S vs. F: Training on slow walk and testing on fast walk.
(b) F vs. S: Training on fast walk and testing on slow walk.

Fig.3 The CMS curves of LDT. (a) Different iteration times; (b) Different latent space dimensions.

The iteration times and latent space dimension are two influence factors on the performance of LDT. The experiment results of S vs. F with different iteration times $n$ (represented as LDT-$n$) and latent space dimension $m$ (represented as LDT($m$)) are shown in Fig.3, respectively. The iteration times are changed from 10 to 50 in Fig.3(a). High iteration times do not always have better performance. 30 iteration times gives the highest recognition rate at rank 1. Although its whole performance is worse than that of 40 iteration times, the performance of rank 1 is more important and iteration times 30 is chosen. Fig.3(b) shows the CMS curves of different latent space dimension. The performance of 5 dimensions performs worst. The dimension is chosen as 20.

The CMS curves of the experiments are illustrated in Fig.4. The performance of DT is the worst in the two experiments, only 32% for experiment S vs. F and 20% for experiment F vs. S at rank 1. Although KDT greatly
improves the performance, it is still 20% worse than LDT at rank 1.

Fig.4 The CMS curves of DT, KDT and LDT. (a) S vs. F; (b) F vs. S.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a logistic dynamic texture model (LDT) to describe binary data sequence. This model is an extension of the dynamic texture model and its parameters are trained with logistic PCA. Logistic PCA processes the pixels of 1 and 0 separately, when decomposing the binary image sequence. Proper iteration times and dimension of latent space can be chosen according to experiments. The experiment results show its superiority. LDT can be further improved from the following aspects. Firstly, logistic PCA randomly chooses the initialization of its parameters. Secondly, a feedback can be added to the model for error correction.

5. REFERENCES


