Gait recognition based on improved dynamic Bayesian networks

Changhong Chen [a,*], Jimin Liang [b], Xiuchang Zhu [a]

[a] Image Processing and Image Communication Lab, College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[b] Life Sciences Research Center, School of Life Sciences and Technology, Xidian University, Xi'an 710071, China

ARTICLE INFO

ABSTRACT

In this paper, we proposed an improved two-level dynamic Bayesian network layered time series model (LTSM), which aims to solve the limitations hindering the application of available dynamic Bayesian networks, the hidden Markov model (HMM) and the dynamic texture (DT) model to gait recognition. In the first level, a gait silhouette or feature cycle is divided into several temporally adjacent clusters. Each cluster is modeled by a DT or logistic DT (LDT). In the second level, HMM is built to describe the relationship among the DTs/LDTs. Besides LTSM, LDT is also an improved dynamic Bayesian network presented in this paper to describe the binary image sequence, which introduces the logistic principle component analysis (PCA) to learning its parameters. We demonstrated the validity of LTSM with experiments on both the CMU Mobo gait database and CASIA gait database (dataset B), and that of LDT on the CMU Mobo gait database. Experimental results showed the superiority of the improved dynamic Bayesian networks.

1. Introduction

Gait recognition aims to identify people by the way they walk. In comparison with other biometrics, gait pattern has the advantages of being unobtrusive, difficult to conceal and effective at a distance. However, the gait recognition algorithm has to deal with the image sequences instead of a single image. Both the spatial information of the gait image and the temporal transformation of the gait sequence are important. Methods characterizing the spatial and temporal information fall into two categories: information fusion and dynamic Bayesian networks.

Information fusion technology offers a promising solution to combine the spatial and temporal information to develop a superior classification system. Wang et al. [1] derived the dynamic information of gait by using a condensation framework to track the walker and to recover joint-angle trajectories of lower limbs. The static body information is derived from temporal pose changes of the segmented moving silhouettes which are represented as an associated sequence of complex vector configurations and are then analyzed using the Procrustes shape analysis method to obtain a compact appearance representation. Both the static and dynamic cues may be used independently for recognition and are also fused on the decision level using different combinations of rules to improve identification and verification performances. Lam et al. [2] presented two gait feature representation methods, the motion silhouette contour templates (MSCTs) and static silhouette templates (SSTs), and performed decision-level fusion by summarizing the similarity scores. Bazin et al. [3] examined the fusion of a dynamic feature and two static features in a probabilistic framework. The dynamic signatures derived from bulk motion and shape characteristics of the subject. The first static signature is a vector yielded from an average silhouette. The second static vector is obtained from a block-based silhouette averaging method. They proposed a process for determining the probabilistic match scores using intra- and inter-class variance models together with the Bayesian rule. Veres et al. [4] proposed a two-stage data fusion rule to fuse the dynamic feature and the first static signature mentioned in Ref. [3]. In summary, the feature fusion method is a good choice to combine spatial and temporal information, but it is challenging to extract the representative dynamic feature, especially when the gait silhouettes are of low quality.

Dynamic Bayesian networks do not need to extract dynamic features and embody the temporal transformation via the model parameters. The most commonly used dynamic Bayesian networks are descendants of either the hidden Markov models (HMM) or stochastic linear dynamical systems, which are also known as state-space models (SSM) [5].

HMM represents information regarding the past of a sequence through a single discrete random variable—the hidden state. Kale et al. [6] and Sundaresan et al. [7] introduced HMM to gait recognition, which consider two different image features: the

* Corresponding author. Tel.: +86 25 83492402; fax: +86 25 83492420.
E-mail address: chenhh@njupt.edu.cn (C. Chen).

width of the outer contour of a binary silhouette and the entire binary silhouette itself. They proposed an indirect approach and a direct approach to train HMM. The indirect approach to forming a feature vector uses a frame to exemplar distance (FED) which captures a subject's shape and his/her motion, under the assumption that the camera is sufficiently distant so that the moving subject can be considered to be planar. The information in the FED vector sequences is captured using an HMM. The direct approach used the feature vector directly (as opposed to computing the FED) for training an HMM. The observation probability is estimated using the distance between the exemplars and the image features. In Ref. [8], the direct approach of an HMM was further developed. The width vector of the outer contour is used as the feature, and an adaptive method is developed to calculate the observation probability of HMM. While the performance of HMM is excellent, there are still some remaining problems that need to be addressed, such as how to determine the hidden states. Liu and Sarkar [9] proposed a population hidden Markov model (pHMM) for silhouette reconstruction and cleaning. The pHMM helps to map a frame in any given sequence to a stance, and an appearance-based Eigen–Stance model is used to reconstruct the computed silhouette in the frame. This method can reconstruct silhouettes that are visually appealing and robust to viewpoint variation, but the lost characteristic possessed by a single subject drops the recognition performance.

SSM represents information of the past through a real-valued hidden state vector. The dynamic texture (DT) model [10], a linear SSM, is an efficient method for dynamic image sequence modeling. DT learns the parameters through a closed-form solution and commonly uses the principal component analysis (PCA) to get the observation parameters. PCA assumes a Gaussian distribution over a set of observations, while some study has shown that a natural way to model binary data is using Bernoulli distribution [11,12]. Therefore, it is not suitable for directly using DT to model gait sequences. Some researches [13–15] applied DT to model the extracted gait features. Mazzaro et al. [13] measured the angles of the shoulder, elbow, hip and knee joints and used the angle vector to get a nominal model. Bissacco et al. [14] introduced a representation with the projection features, which encode the distance of the points on the silhouette from lines passing through its center of mass. A linear non-Gaussian model was constructed with the features. The autoregressive and moving average (ARMA) model in Ref. [15] is also a DT, which is constructed on the tangent space projections of a shape sequence extracted from the binary images by uniform sampling or uniform arc-length sampling. Bissacco and Soatto [16] proposed a hybrid autoregressive model of human motion and novel algorithms to estimate switches and model parameters. A set of joint-angle trajectories on the skeletal model were extracted to train the hybrid model for dynamic discrimination. Two problems exist in the aforementioned applications. First, these models are good at describing a motion process, but they are not accurate in detailed information. Therefore, these methods are only validated based on activity recognition of a small database or classifying gait styles. Second, the gait sequence is a non-linear process, while the DT and its extensions are linear models.

Chan and Vasconcelos [17–19] further developed the DT method. They improved the modeling capability of DT by introducing kernel PCA to learn a non-linear observation function [17]. In Ref. [18], the mixture of DTs was studied. It is a statistical model for an ensemble of video sequences that is sampled from a finite collection of visual processes, each of which is a DT. The mixture of DTs is shown to be a suitable representation for both the appearances and dynamics. A novel video representation, the layered DT is further discussed in Ref. [19]. It represents a video as a collection of stochastic layers of different appearances and dynamics. Each layer is modeled as a temporal texture sampled from a different linear dynamic system. These extensions are good at describing dynamic

textures, but are not suitable for gait sequencing. Furthermore, it is not needed to build complicated models such as [18,19] for binary gait sequences.

In this paper, an improved dynamic Bayesian network, the logistic DT (LDT), is proposed to directly model the binary image sequence. It introduces logistic PCA to learn the observation function, which assumes a Bernoulli distribution over a set of observations and processes the pixels of 1 and 0 separately. This model can avoid the loss of useful information caused by feature extraction. It is evaluated using the CMU Mobo gait database. It is testified to be more suitable for describing a binary image sequence than DT, although it is still a non-linear model.

In order to tackle the limitations hindering the application of HMM and DT/LDT to gait recognition, the major contribution of this paper is to propose another improved dynamic Bayesian network, the layered time series model (LTSM). An LTSM is a two-level model which combines HMM and DT/LDT. The first level has multiple DTs/LDTs. A gait silhouette or feature cycle is segmented into several temporally adjacent clusters, which can be considered as linear processes. Each cluster can be modeled by a DT/LDT. The second level is an HMM, which is used to describe the statistical distribution of different DTs/LDTs. The DTs/LDTs are treated as the hidden states of the HMM and the observation probability is a function of the distance between the observation and the synthesized observation of DT/LDT. An LTSM conquers the non-linear process representation problem using piecewise linear DTs/LDTs, and then applies HMM to describe the transition of the DTs/LDTs. Its validity is evaluated by both the CMU Mobo gait database [20] and CASIA gait database (dataset B) [21]. The experimental results show its superiority over HMM and DT.

The remainder of this paper is structured as follows. Section 2 proposes the LDT in detail. The construction of LTSM is described in Section 3. In Section 4, we evaluate LDT and LTSM using the CMU Mobo gait database and CASIA gait database (dataset B). We discuss the results and conclude this paper in Section 5.

2. Logistic dynamic texture (LDT) model

In this section, we simply introduce DT and then proposed an improved dynamic Bayesian network LDT to model the binary image sequence.

2.1. DT

DT is a generative stochastic model that treats the video sequence as a sample from the output of a linear SSM. It was first proposed by Doretto [10], and it was a successful application of the linear dynamical system to video processing. 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 visual component $y_t \in R^D$ is a function of the current state
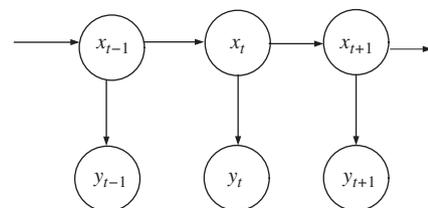

Fig. 1. Topology structure of dynamic texture model (DT).