International Journal of Digital Information and Wireless Communications (IJDIWC) 1(2): 466-515
The Society of Digital Information and Wireless Communications, 2012(ISSN 2225-658X)

Gait Analysis and Recognition Using Multi-Views Gait Database

Tee Connie[1], Michael Kah Ong Goh[1], Andrew Beng Jin Teoh[2]
[1] Multimedia University, Jalan Ayer Keroh Lama, 75450, Melaka, Malaysia
{tee.connie, michael.goh}@mmu.edu.my
[2] Electrical and Electronic Engineering Department, Yonsei University, Seoul, South Korea
bjteoh@yonsei.ac.kr

ABSTRACT

This paper presents an automatic gait recognition system that recognizes a person by the way he/she walks. The gait signature is obtained based on the contour width information of the silhouette. Using this statistical shape information, we could capture the compact structural and dynamic features of the walking pattern. As the extracted contour width feature is large in size, Fisher Discriminant Analysis is used to reduce the dimension of the feature set. After that, a modified Probabilistic Neural Networks is deployed to classify the reduced feature set. Satisfactory result could be achieved when we fuse gait images from multiple viewing angles. In this paper, we aim to identify the complete gait cycle of each subjects. Every person walks at different paces and thus different number of frame sizes are required to record the walking pattern. As such, it is not robust and feasible if we take a fixed number of video frames to process the gait sequences for all subjects. We endeavor to find an efficient method to identify the complete gait cycle of each individual. Towards this end, we can work on succinct representation of the gait pattern which is invariant to walking speed for each individual.

KEYWORDS

Gait recognition, Statistical shape analysis, Fisher Discriminant Analysis, Probabilistic Neural Networks.

1 INTRODUCTION

Recently, gait recognition has emerged as an attractive biometric technology to identify a person at a distance. Gait recognition is used to signify the identity of a person based on the way the person walks [1]. This is an interesting property by which to recognize a person, especially in surveillance or forensic applications where other biometrics may be inoperable. For example in a bank robbery, it is not possible to obtain face or fingerprint impressions when masks or hand gloves are worn. Therefore, gait appears as a unique biometric feature to allow possible tracking of people's identities.

Gait has a number of advantages as compared to the other biometric characteristics. Firstly, gait is unique to each individual. Every person has a distinctive way of walking due to the different biomechanical and biological compositions of the body [2]. Human locomotion is a complex action which involves coordinated movements of the limbs, torso, joints, and interaction among them. The variations in body structures like height, girth, and skeletal dimension can also provide cue for personal recognition. Secondly, gait is unobtrusive. Unlike other biometrics like fingerprint or retina scans which require careful and close contact with the sensor, gait recognition does not require much cooperation from the users. This unobtrusive nature makes it suitable for wide range of surveillance and security applications. Thirdly, gait can be used for recognition at a distance. Most of the biometrics such as iris, face, and fingerprint require medium to high quality images to obtain promising recognition performance. However, these good quality images can only be acquired when the users are standing close to the sensors or with specialized sensing hardware. When these controlled environments are not applicable, like in most of the surveillance systems in real-life, these biometrics features are rendered of little use. Therefore, gait appears as an attractive solution because gait is discernable even from a great distance. Lastly, gait is difficult to disguise or conceal. In many personal identification scenarios, especially those involving serious crimes, many prominent biometric features are obscured. For instance, the face may be hidden and the hand is obscured. However, people need to walk so their gait is apparent. Attempt to disguise the way a person walks will make it appears even more awkward. In a crime scene, criminals would want to leave at speed and would not want to provoke attention to minimize the chance of capture. Therefore, it is extremely hard for the criminals to masquerade the way they walk at the crime scene without drawing attention to themselves. These wealth of advantages make gait recognition a graceful technology to complement the existing biometric applications.

There are two main approaches for gait recognition, namely model-based [3], [5] and appearance-based [2], [6], [7], [8]. The model-based approach explicitly models the human body based on body parts such as foot, torso, hand, and leg. Model matching is usually performed in each frame to measure the shape or dynamics parameters. Cunado et al. [3] assumed legs as an interlinked pendulum, and gait signature was derived from the thigh joint trajectories as frequency signals. Johnson and Bobick [4] used activity-specific static body parameters for gait recognition without directly analyzing gait dynamic. Yam et al. [5] recognized people based on walking and running sequences and explored the relationship between the movements that were expressed as a mapping based on phase modulation. Usually, the model-based approaches are easy to be understood. However, these methods require high computational cost due to the complex matching and searching processes involved.

On the contrary, the appearance-based approach is more straight-forward. These methods generally apply some statistical theories to characterize the entire motion pattern using a compact representation without considering the underlying motion structure. BenAbdelkader et al. [2] obtained eigengaits by using image self-similarity plots. Lu and Zhang [6] exploited Independent Component Analysis (ICA) to extract gait feature from human silhouettes. Tao et al. [7] also applied similar approach to represent the gait sequences using Gabor gait and Tensor gait features. On the other hand, Kale et al. [8] modeled the temporal state-transition nature of gait by using Hidden Markov Model (HMM). The appearance-based approaches are more often used due to their simpler implementation process and lower computational cost.

In this paper, we adopt the appearance-based approach to keep computational complexity low. Our method exploits a less complex, yet effective approach to analyze and recognize the gait feature. We first extract the binary silhouettes from the gait sequences. The reason we work on silhouette images is because silhouette images are invariant to changes in clothing color/texture and also lighting condition. After that, we extract contour width feature from the silhouette images. Fisher Discriminant Analysis (FDA) is then applied to reduce the dimension of this feature set. Next we deploy a modified PNN as the classifier. As the gait sequences we used are composed of multi-view angles datasets [9], the scores obtained from each viewing angles are fused to obtain the final result. The overall framework of the proposed system is depicted in Figure 1.

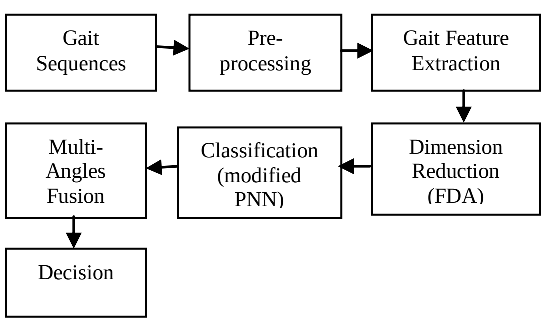


Figure 1. Framework of the proposed system.

The contributions of this paper are two-fold. Firstly, we modify the pattern layer of PNN in order to characterize the gait signatures more effectively. Secondly, we analyze the motion model of the human gait in order to determine the walking cycle of each person. By learning the walking cycles, we could find a compact representation of the walking pattern for each individual. Figure 2 illustrates the five stances corresponding to one walking cycle of a subject. Being able to identify the walking cycle is important because every person walks at different paces. Some people transit between the successive stances very quickly as they walk very fast, and vice versa. If we fix a number of video frames for analysis for every person, the video frames for the person who is walking fast may contain repeating walking pattern. On the other hand, the video frames for the person who is walking slowly will not be able to capture the entire walking pattern for that person. Therefore, we endeavor to identify the gait cycle of each subject (which may encompass varying number of video frames), and use this compact representation to process the gait motion more precisely.

2 PROPOSED SYSTEM

2.1 Pre-Processing

Given a video sequence, we use the background subtraction technique [10] to obtain the silhouette of the subject. The background-subtracted silhouette images may contain holes, noise or shadow elements. Therefore, we apply some morphological operations like erosion and dilation to fill the holes and remove the noises. A binary connected component analysis is then used to obtain the largest connected region which is the human silhouette in the image.

After obtaining the silhouette images, the gait cycle for each subject could be identified by measuring the separation point of the feet. The walking cycle of a person comprises a generic periodic pattern, with an associated frequency spectrum [3]. Let say the walk starts by lifting the right foot and moving forward. When the right foot strikes the floor ("heel-strike"), the left leg flexes and lifts from the ground ("heel-off"). The body moves by interchanging the movement between the left and right foot lifting and striking the floor, forming a periodic gait cycle (Figure 2).

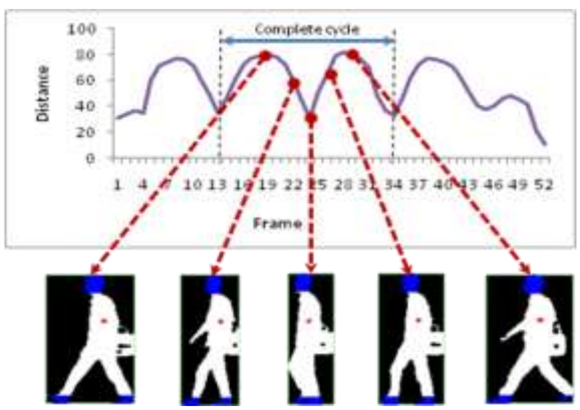


Figure 2. Five stances corresponding to a gait cycle.

A complete gait cycle could be determined by analyzing the gait signature graph and locating the local minimal of the graph (Figure 2). The gait signature is constructed based on separation between the two legs. The start of the gait cycle is signified by the second local minima point detected in the graph. We do not consider the first local minima because it may contain some erroneous "heel-strike" or "heel-off" events. The complete gait cycle could be extracted by taking two consecutive temporal transitions between the "heel-strike" and "heel-off" phases as indicated by the positions of two successive local minima points in the gait signature.

The gait cycle for each person comprises different lengths depending on the speed the person walks. The frame sequences to capture a slow walking pattern may be longer than that of a fast walking pattern. We analyzed the frame sequences of all subjects in our dataset and found that the average number of frames to characterize a complete gait cycle is about 20 frames. Therefore, we take this number to optimally represent the gait cycle for the subjects in our experiment. For some slow-moving subjects with longer frame sequences, we "interpolate" the frame sequence length by taking alternative walking poses to reduce the number of frame sequences. For fast-moving subjects whose frame numbers are less than 20, we "extrapolate" the frame sequence by taking a few more frames beyond the end of the walking cycle to make up the figure. The frame sequences containing the gait cycle is then extracted from the whole video sequence for further processing. Instead of having to work on lengthy frame sequences containing repeating walking patterns, we can focus on a compact representation of the gait movement for analysis.

2.1 Gait Feature Extraction

We obtain the width features of the contour along each row of the image and store them as the feature set. Assume that F denotes the number of frames in a gait sequence, and R refers to the number of rows encompassing the subject contour. The sequence of width vector of the gait cycle of a subject can be represented by $X = \{x_1^{w_i}, x_2^{w_i}, ..., x_F^{w_i}\}$, where w_i refers to the vector corresponding to each row in an image, and i = 1, 2, …, R. X is the result of stacking the contour width vector

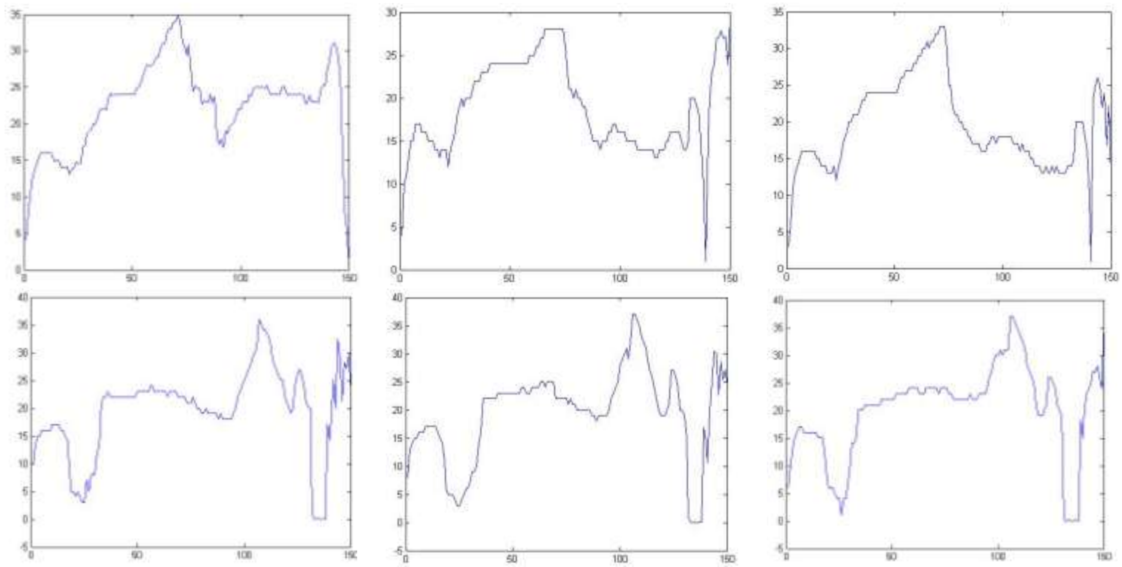


Figure 3. The extracted contour width features from two subjects (depicted in different rows).

together to form a spatio-temporal pattern. Some sample contour width features extracted from two different subjects are depicted in Figure 3. We observe that the contour width features portray little differences for the same subject while significant variations among different subjects. Therefore, the contour width feature qualifies as a gait feature to distinguish individuals.

One advantage of representing the gait signature using the contour shape information is that we do not need to consider the underlying dynamics of the walking motion. It is sometimes difficult to compute the joint angle, for instance, due to self-occlusion of limbs and joint angle singularities. Therefore, the contour/shape representation enable us to study the gait sequence from a holistic point of view by implicitly characterizing the structural statistics of the spatio-temporal patterns generated by the silhouette of the walking person. Note that X is a high-dimensional vector and modeling this feature set requires a lot computational cost. Hence, some dimension reduction technique is used to minimize the size of this feature set.

2.3 Dimension Reduction using FDA

FDA is a popular dimension reduction technique in the pattern recognition field [11]. FDA maximizes the ratio of between-class scatter to that of within-class scatter. In other words, it projects images such that images of the same class are close to each other while images of different classes are far apart. Given the contour width feature vector, $\{X_1^1, X_2^1, ..., X_n^1, X_1^2, X_2^2, ..., X_n^C\}$ where c denotes the number of classes and n represents the total number of gait samples per class. Let the mean of images in each class and the total mean of all images be represented by m_c and m, respectively, the images in each class are centered as,

$$\phi_n^c = X_n^c - m_c \quad (1)$$

and the class mean is centered as,

$$\omega_c = m_c - m \qquad (2)$$

The centered images are then combined side by side into a data matrix. By using this data matrix, an orthonormal basis U is obtained by calculating the full set of eigenvectors of the covariance matrix $\phi_n^{cT}\phi_n^c$. The centered images are then projected into this orthonormal basis as follow,

$$\phi_n^c = U^T \phi_n^c \qquad (3)$$

The centered means are also projected into the orthonormal basis as,

$$\omega_c = U^T \omega_c \qquad (4)$$

Based on this information, the within class scatter matrix S_W is calculated as,

$$S_W = \sum_{j=1}^{c}\sum_{k=1}^{n_j} \phi_k^j \phi_k^{j\,T} \qquad (5)$$

And the between class scatter matrix S_B is calculated as,

$$S_B = \sum_{j=1}^{C} n_j \omega_j \omega_j^T \qquad (6)$$

The generalized eigenvectors, V, and eigenvalues, λ, of the within class and between class scatter matrix are solved as follow,

$$S_B V = \lambda S_W V \qquad (7)$$

The eigenvectors are sorted according to their associated eigenvalues. The first $M-1$ eigenvectors are kept as the Fisher basis vectors, W. The rotated images, α_M where $\alpha_M = U^T i_M$ are projected into the Fisher basis by

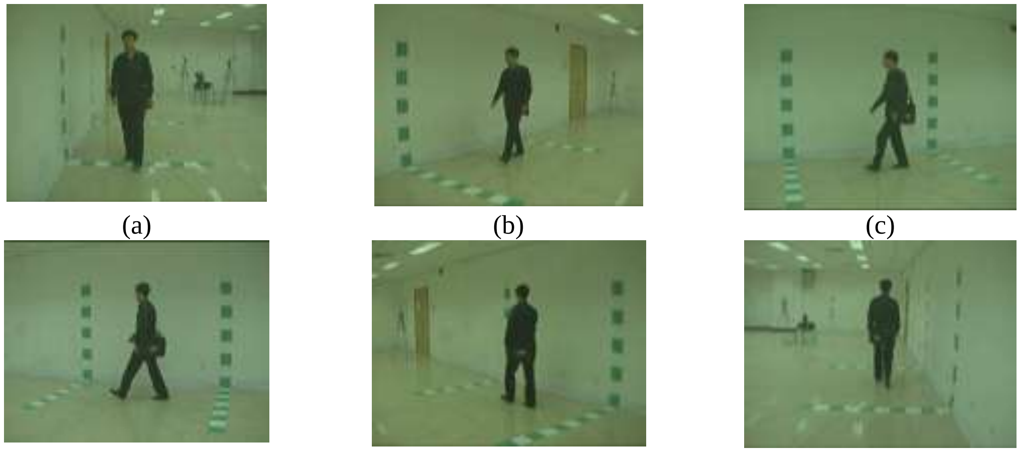
$$\varpi_{nk} = W^T \alpha_M \qquad (8)$$

where $n = 1, \ldots, M$ and $k=1, \ldots, M-1$.

The weights obtained is used to form a vector $\chi = [\varpi_{n1}, \varpi_{n2}, \ldots, \varpi_{nK}]$ that describes the contribution of each fisher basis in representing the input image. In this paper, the original dimension of X is R x F. This size can be reduced to only c-1 after FDA processing.

2.4 Classification using Modified PNN

We modify an ordinary PNN to classify the gait features. In general, a PNN consists of three layers – a pattern, summation and output layers (apart from the input layer) [12]. The pattern layer contains one neuron for each input vector in the training set, while the summation layer contains one neuron for each user class to be recognized. The output layer merely holds the maximum value of the summation neurons to yield the final outcome (probability score). The network can simply be established by setting the weights of the network using the training set. The modifiable weights of the first layer are set by $\omega_{ij} = x^t_{ij}$ where ω_{ij} denoting the weight between ith neuron of the input layer and jth neuron in the pattern layer, and x^t_{ij} is the ith element of the contour width feature, X, j in training set. The second layer weights are set by $\omega_{jk} = T_{jk}$, where ω_{jk} is the weight between neuron j in pattern layer and neuron k of the output layer, and 1 is assigned to T_{jk} if pattern j of the training set belongs to user k and 0 otherwise.



(a) (b) (c) (d) (e) (f)

Figure 5. Sample walking sequences from different viewing angles at, (a) 0°, (b) 36°, (c) 72°, (d) 108°, (e) 144°, (f) 180°.

(a) (b) (c)

Figure 4. The subject walks under different conditions. (a) Walking normally, (b) Walking with a coat, (c) Walking with a bag.

After the network is trained, it can be used for classification task. In this paper, the outcome of the pattern layer is changed to the form expressed in Equation 9 instead of inner product that is used in standard PNN.

$$out_{\boldsymbol{j}} = \exp\left(-\left(\sum_{\boldsymbol{i}=1}^{m}(\boldsymbol{x}_i \text{-} \omega_{ij})\right)/\sigma\right) \qquad (9)$$

Note that out_{j} is the output of neuron j in pattern layer, and x_i refers to ith element of the input. σ is the smoothing parameter of the Gaussian kernel which is the only independent parameter that can be decided by the user. The input of the summation layer is calculated by the equation

$$in_k = \sum_{j=1}^{n} out_j \times \omega_{jk} \qquad (10)$$

where in_k is the input of neuron k in output layer. The outputs of the summation layer are binary values, i.e 1 is assigned to out_k if in_k is larger than the input of others neurons and 0 otherwise.

In the experiment, we are using cross-validation to estimate the accuracy of the method more reliably. The smoothing parameters (σ_1, σ_2,…, and σ_j) need to be carefully determined in order to obtain an optimal network. For convenience sake, a straightforward procedure is used to select the best value for σ. Firstly, an arbitrary value of σ is chosen to train the network, and then test it on a test set.

This procedure is repeated for other σ's values and the σ giving the least errors is selected. In this paper, 0.1 appears to be best σ. The motivation of using the modified PNN is driven by its ability to better characterize the gait features and its generalization property. Besides, PNN only requires one epoch of training which is good for online application.

3 EXPERIMENTAL RESULTS

3.1 Experiment Setup

In this paper, we use the publicly available CASIA gait database: Database B [9]. The gait data in this database consists of views from eleven different angles (Figure 4). Besides, the database also contains subjects walking under different conditions like walking with coats or bags (Figure 5). There are ten walking sequences for each subject, with six samples containing subjects walking under normal condition, two samples with subjects walking with coats, and two samples with subjects carrying bags.

We selected the first fifty subjects in the database to be used in this paper. Among the ten gait sequences for each subject, we used three samples under the normal walking condition as gallery set. The remaining seven samples under the normal walking condition (three samples), walking with coats (two samples) and bags (two samples) are used as the probe sets.

To consolidate the gait sequences for the different viewing angles, the sum-rule based fusion rule is adopted in this paper. This fusion method is selected because of its good performance as compared to AND- and OR-fusion rules, or even more sophisticated techniques like neural networks [13] and decision trees [14].

3.2 Verification Performance under Different Viewing Angles

We have conducted a number of experiments to testify the performance of the proposed method. The four important biometric performance measurements criterion namely False Rejection Rate (FRR), False Acceptance Rate (FAR), Equal Error Rate (EER), and Genuine Acceptance Rate (GAR) are used to evaluate the performance of the system. FRR refers to the percentage of clients or authorized person that the biometric system fails to accept. FRR is defined as

$$\mathrm{FRR} = \frac{N_{RC}}{N_C} \times 100\% \tag{11}$$

where N_{RC} refers to the number of rejected clients and N_C denotes the total number of client access. On the other hand, FAR represents the percentage of imposters or unauthorized person that the biometric system fails to reject. FAR is defined as

$$\mathrm{FAR} = \frac{N_{AI}}{N_I} \times 100\% \tag{12}$$

where N_{AI} signifies the number of accepted imposters and N_I represents the total number of imposter access. EER is an error rate where FAR is equals to FRR. It is commonly used to determine the overall accuracy of the system and it serves as a comparative measure against the other biometric systems. On the other hand, GAR denotes the percentage of clients or authorized person that the biometric system correctly accepts. GAR