Review on Vision-Based Gait Recognition: Representations, Classification Schemes and Datasets

Chin Poo Lee, Alan Wee Chiat Tan and Kian Ming Lim

Faculty of Information Science and Technology, Multimedia University, Malacca, Malaysia

Abstract: Gait has unique advantage at a distance when other biometrics cannot be used since they are at too low resolution or obscured, as commonly observed in visual surveillance systems. This paper provides a survey of the technical advancements in vision-based gait recognition. A wide range of publications are discussed in this survey embracing different perspectives of the research in this area, including gait feature extraction, classification schemes and standard gait databases. There are two major groups of the state-of-the-art techniques in characterizing gait: Model-based and motion-free. The model-based approach obtains a set of body or motion parameters via human body or motion modeling. The model-free approach, on the other hand, derives a description of the motion without assuming any model. Each major category is further organized into several subcategories based on the nature of gait representation. In addition, some widely used classification schemes and benchmark databases for evaluating performance are also discussed.

Keywords: Gait, Gait Recognition, Gait Datasets

Introduction

In recent decades, much research effort has been devoted to the study of vision-based gait recognition. The aim of gait recognition is to automatically describe walking pattern from video sequences and thereafter identify individual based on the walking pattern. The basic architecture of a vision-based gait recognition system is depicted in Fig. 1.

The second stage of the system is feature extraction, where gait sequences are mapped into a compact set of gait features, or gait signatures. There are many candidate methods for this task and they can be broadly grouped into two major categories, i.e., model-based and model-free. In the former, a set of body or motion parameters is obtained via human body or motion modeling. The model-free approach, on the other hand, derives a description of the motion without assuming any model. The set of features of the known classes obtained in this stage serves as the gait templates and stored in the gait library. A review of model-based and model-free approaches are presented in section 2.1 and 2.2, respectively.

The last stage of the system is the pattern recognition. The aim here is to identify, given the observed gait signatures of an unknown class, the optimal match from the library of known classes. A review of the commonly used pattern classification schemes is reported in section 3.

Feature Extraction

This section briefly discusses some of the most popular feature extraction methods in the literature. The feature extraction methods can be broadly divided into two major categories, i.e., model-based and model-free. For each of the categories, there are some subcategories, as depicted in Fig. 2.

Fig. 1. The framework of vision-based gait recognition
Model-Based Approaches

Model-based approaches describe walking pattern using model parameters of human body components or motion, such as motion trajectories, limb lengths, limb angular speeds and etc. Two commonly used gait representation in model-based approaches are structural model and motion model.

Structural Model

A structural model is a model that describes the properties of body components via measurements of the limb lengths, distance between limbs and relative position of limbs, among others. A structural model can be made up by approximating human body using primitive shapes, stick figures, or arbitrary shapes (Yam and Nixon, 2009).

An early work by Bobick and Johnson (2001) measured static body parameters when the feet are maximally apart. Despite the simplicity in calculating static body parameters, these features suffer from the loss of deformation data. For this reason, researchers often turn to establishing and measuring the structural model throughout the gait cycle.

BenAbdelkader et al. (2002) obtained a stick figure from the image by locating the silhouette bounding box and the mid-feet point, as depicted in Fig. 3. From the stick figure, they estimated the stride parameters (speed, cadence and stride length) and the height parameters of the moving silhouette. Later, Zhang et al. (2004) employed a two-dimensional five-link body model to represent the walking pattern. They extracted the gait features from image sequences using Metropolis-Hasting method. Likewise, Yoo and Nixon (2011) represented silhouette by a planar stick figure with eight sticks and six joints, as shown in Fig. 4. Motion parameters are then measured from the trajectories of the stick figure.

Another variant is the work by Lee and Grimson (2002), where silhouette was separated into seven ellipsoidal regions. From these ellipses, they derived moment-based features including centroid, aspect ratio of the major and minor axes, as well as the orientation of major axis of the ellipse. They also measured the Fourier transform of ellipse parameters in temporal axis. Figure 5 displays the ellipsoidal model. Wagg and Nixon (2004) proposed an articulated model in which ellipses were fitted for the torso and the head, lines for the legs and a rectangle for each foot, as depicted in Fig. 6.

Huang and Boulgouris (2009), the silhouettes were manually labeled into eight body components. Some geometry features, i.e., the area, the center of gravity and the orientation were thereafter measured on each body component. In the more recent work by Tafazzoli and Safabakhsh (2010), the silhouette was segmented into three major regions, i.e., the head, torso and leg, based on the mean anatomical proportions. Then, active contour models and Hough transform were used to construct a posterior model of the gait motion. Fourier analysis was subsequently used to reveal the motion patterns of the body parts. Table 1 summarizes the structural model gait representation.

Motion Model

A motion model measures the parameters of gait mechanics, such as the kinematics of joint angles in human walking.

In the work of Cunado et al. (1997; 2003), the movement of the legs was fitted into a pendulum-like motion model. Fourier transform analysis, thereafter, was used to describe the frequency components of the leg movements. Similarly, Yam et al. (2004) modeled the lower limbs as double pendulum; the angular motion was thereafter recorded as the phase-weighted magnitude of the Fourier descriptor of the lower limbs.

Tanawongsuwan and Bobick (2001) placed markers on legs and thorax to derive the joint angle trajectories. The variation in joint angles in temporal axis was then derived. Another variant of the motion model was presented in Yoo et al. (2002) where hip and knee angles were estimated from the silhouette by linear regression analysis. The gait signature was denoted as the parameters obtained from the trigonometric-polynomial interpolant functions of the angles. In Wang et al. (2004), a Conditional Density Propagation framework (Isard and Blake, 1998) was used to track the human and to further estimate joint angle trajectories of the lower limbs. The angles of joints were represented as Euler angles.
In a later work, Fathima and Banu (2012) performed skeletonisation on the extracted silhouette of each image frame. Then six joint angles from head to foot were calculated. More recently, Lu et al. (2014) extracted motion angles of lower limbs to build joint distribution spectrums. Based on the joint distribution, the feature histogram was thereafter computed as gait signature. A summary of motion model is provided in Table 2.

Fig. 3. The stick figure model in BenAbdelkader et al. (2002)

Fig. 4. The stick figure model in Yoo and Nixon (2011)

Fig. 5. The ellipsoidal model in Lee and Grimson (2002)

Fig. 6. The articulated model in Wagg and Nixon (2004)

Fig. 7. The first column shows a sample frame of the actions. The second and third columns show their corresponding MEIs and MHIs (Bobick and Davis, 2001)

Though model-based approaches are more robust to view and scale variations, accurately locate the joints positions is a strenuous task due to the non-rigid structures of the human body and to self-occlusion (Yang et al., 2008; Wang et al., 2011). For this reason, researchers often turn to model-free approaches.

Model-Free Approaches

Model-free approaches directly extract motion pattern from the gait sequences without constructing any model. The gait representation obtained by model-free approaches can be broadly categorized into appearance-based representation, transformation-based representation and distribution-based representation.
Table 1. Summary of model-based approaches (structural model)

<table>
<thead>
<tr>
<th>Literature</th>
<th>Gait features</th>
<th>Classifier/Distance metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobick and Johnson (2001)</td>
<td>Static body parameters</td>
<td>Population covariance</td>
</tr>
<tr>
<td>BenAbdelkader et al. (2002)</td>
<td>Stick Fig. (stride and height parameters)</td>
<td>Bayesian classifier</td>
</tr>
<tr>
<td>Zhang et al. (2004)</td>
<td>Five-link biped (linked feature trajectories)</td>
<td>HMMs</td>
</tr>
<tr>
<td>Yoo and Nixon (2011)</td>
<td>Stick Fig.</td>
<td>kNN</td>
</tr>
<tr>
<td>Lee and Grimsson (2002)</td>
<td>Ellipsoidal model (moment-based features)</td>
<td>kNN + Mahalanobis distance</td>
</tr>
<tr>
<td>Huang and Boulgouris (2009)</td>
<td>8 body components (geometry features)</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Tafazzoli and Safabakhsh (2010)</td>
<td>3 body regions (posterior model of motion)</td>
<td>kNN</td>
</tr>
</tbody>
</table>

Table 2. Summary of model-based approaches (motion model)

<table>
<thead>
<tr>
<th>Literature</th>
<th>Gait features</th>
<th>Classifier/distance metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cunado et al. (1997; 2003)</td>
<td>Pendulum (inclination of legs)</td>
<td>KNN</td>
</tr>
<tr>
<td>Yam et al. (2004)</td>
<td>Pendulum (joint angle trajectories)</td>
<td>kNN + Euclidean distance</td>
</tr>
<tr>
<td>Tanawongsuwan and Bobick (2001)</td>
<td>Joint angle trajectories</td>
<td>DTW</td>
</tr>
<tr>
<td>Yoo et al. (2002)</td>
<td>Joint angle trajectories</td>
<td>Neural networks</td>
</tr>
<tr>
<td>Wang et al. (2004)</td>
<td>Joint angle trajectories</td>
<td>kNN + Euclidean distance</td>
</tr>
<tr>
<td>Pathima and Banu (2012)</td>
<td>Joint angle trajectories</td>
<td>SVM</td>
</tr>
<tr>
<td>Lu et al. (2014)</td>
<td>Joint angle trajectories</td>
<td>SVM</td>
</tr>
</tbody>
</table>

**Appearance-Based Representation**

In appearance-based representation, the gait motion is accumulated into an energy image. The higher the energy at the position in the image, the more frequent the motion occurs at the position.

From a sequence of gait images, Bobick and Davis (2001) obtained two temporal templates, namely Motion Energy Image (MEI) and Motion History Image (MHI). The former is a binary image representing the location of motion in an image sequence. MHI, on the other hand, is a grayscale image representing the recency of motion. Figure 7 shows a few actions and their corresponding MEIs and MHIIs. Although MHI attempts to capture the direction of motion, it suffers from several drawbacks. A major problem of MHI method lies in its difficulty to discriminate the motion direction when there is self-occlusion (Ahad et al., 2012). In addition to that, the MHI also has the drawback that it is sensitive to the variance of motion duration. In view of this, Lee et al. (2014b) proposed a time-sliced averaged motion history image (TAMHI). In their work, the gait cycle is divided into several regular time windows to generate multi-composite images to better preserve transient information. Histograms of Oriented Gradients (HOG) descriptors are then calculated on these composite images to obtain the gait signature. Figure 8 shows the TAMHI composite images and the TAMHI-HOG descriptors of each time window in the gait cycle.

Liu and Sarkar (2004) proposed an averaged silhouette approach, as displayed in Fig. 9. They aligned and averaged the silhouettes to describe the normalized accumulative energy in every gait cycle. An extension of the averaged silhouette approach was endeavored by Xu et al. (2006). In their work, the binary silhouettes were averaged over each gait cycle of the gait sequence. Subsequently, the Coupled Subspace Analysis (CSA) was employed as a preprocessing step to remove noise and, a Discriminant Analysis with Tensor Representation (DATER) was applied to enhance the discriminative power.

Like the averaged silhouette approach, Han and Bhanu (2006) proposed a Gait Energy Image (GEI) representation to denote the averaged accumulative energy of human walking image sequences. They then used a statistical approach to create synthetic templates from real templates, both of which were thereafter fused as the gait signature. Figure 10 depicts the real and synthetic GEI templates.

Some other variations based on GEI had also been introduced. Yang et al. (2008) extracted the dynamic region in GEI using variation analysis. Subsequently, a dynamics weight mask was constructed to intensify the contrast between dynamic region and other regions. The so obtained gait representation, referred to as the Enhanced GEI (EGEI), is shown in Fig. 11. Huang et al. (2013) devised another approach by modifying GEI to extract more information from the lower part of body. A gait representation is devised by convolving the modified GEI with Gabor wavelets. Another variant was proposed in Zhang et al. (2009), where the dynamic variance parts of the GEI were captured in a Dynamic Gait Energy Image (DGEI). Subsequently, they projected the DGEI onto a low-dimensional manifold based on principal component analysis and locality preserving projection (He and Niyogi, 2003). A blend of GEI and fuzzy principal component analysis was presented in Xu and Zhang (2010). They projected the GEI features onto a lower-dimensional space based on fuzzy principal component analysis. Moustakas et al. (2010) proposed another extension, where the gait signatures were obtained via the Radial Integration Transform (RIT) on the GEI and the sequence of silhouettes. Recently, Xu et al. (2012) represented each GEI as a set of local Gabor features.
They combined the Gabor features of different orientations and scales. A global Gaussian Mixture Model (GMM), thereafter, was learnt from the local augmented Gabor features of the entire gallery GEIs. Hu (2014), they represented each GEI as a set of Dual-Tree Complex Wavelet Transform (DTCWT) features. Recently, Choudhury and Tjahjadi (2015) computed the entropy of GEI and subsequently multiscale shape analysis was performed using Gaussian filter.

Inspired by the idea of cumulative energy image, Zhang et al. (2010a) constructed an Active Energy Image (AEI) by accumulating the frame difference between two successive images. Subsequently, each AEI was projected onto a subspace via two-Dimensional Locality Preserving Projections (2DLPP) method. Recently, Huang and Boulgouris (2012) divided the silhouette into three areas, i.e., head, torso and legs. They obtained Shifted Energy Image (SEI) by aligning each area according to their respective horizontal center. Subsequently, the Linear Discriminant Analysis (LDA) was performed on the SEI for dimension reduction. Unlike GEI which accumulates all the images in a gait cycle, Chen et al. (2009) divided a gait cycle into several clusters. Then a Dominant Energy Image (DEI) is produced from each cluster. Summing the cluster’s DEI and the positive portion of the frame difference between consecutive frames, thereafter, produced the Frame Difference Energy Image (FDEI) of the frame. In a more recent development, Roy et al. (2012) introduced a Pose Energy Image (PEI), where they averaged the silhouettes of all key poses in a gait cycle. In addition, the duration spent in each key pose state over a gait cycle was also recorded as the Pose Kinematics. Figure 12 depicts the key poses and their PEIs. A summary of appearance-based representation is presented in Table 3.

Since human walking is a continuous movement of body parts, capturing the directional motion over time is intuitively prominent. In most of the appearance-based representations, there is no explicit consideration of the direction of image motion. Representing motion information in composite energy image, henceforth, leading to the loss of essential directional motion information.

**Transformation-Based Representation**

Besides the normalized accumulative energy approach, transformation is another dominant method to obtain discriminative gait signatures. Some of the widely used transformation methods are Principal Component Analysis (PCA) and Fourier transform.

In an early paper, Murase and Sakai (1996) projected the binarized gait silhouettes onto eigenspace using PCA. Each motion sequence formed a trajectory in the eigenspace, referred to as the parametric eigenspace representation. An extension was devised by Huang et al. (1999), where the gait silhouette images were projected onto a low-dimensional eigenspace based on PCA. The vector obtained from the PCA computation was further projected onto a canonical space based on Canonical Analysis.
Fig. 12. (a) A sequence of silhouette in a gait cycle with the key pose labeled at the bottom of each silhouette. There is a total of 16 key poses, i.e., S1-S16. (b) The PEIs of the 16 key poses (Roy et al., 2012)

Fig. 13. (a) An example of PMS. (b) Computation of shape context for a PMS using log-polar histogram bins (Zhang et al., 2010b)

Another notable use of PCA was presented in Wang et al. (2003a; 2003b). They transformed the silhouette boundaries onto eigenspace using Procrustes shape analysis to obtain the Procrustes Mean Shape (PMS). Accordingly, a Procrustes mean shape distance was proposed as the distance metric. Zhang et al.
(2010b) adopted PMS as their gait signature. They engaged shape context (Belongie et al., 2002) descriptor to measure the similarity between two PMSs. Figure 13 displays a sample PMS and the computation of shape context. Zheng et al. (2011) adopted Partial Least Square regression on the GEI vector to generate optimal feature vector. Subsequently, a robust View Transformation Model (VTM) is obtained by applying robust PCA on the optimal feature vector. Later on, Kusakunniran et al. (2011a) proposed a variant of the Procrustes shape analysis by introducing Pairwise Shape Configuration (PSC) as the shape descriptor. PSC embeds local relation between a boundary point and its neighboring points. Later, they proposed a Higher-order Shape Configuration (HSC) to generate speed-invariant gait features based on Procrustes shape analysis (Kusakunniran et al., 2011b).

Many Fourier descriptor-based techniques can be found in the literature. An early use of Fourier descriptors to model the human gait motion was found in Mowbray and Nixon (2003). The researchers performed Fourier transform on the deformation silhouette boundaries, with the coefficients of the Fourier series being the gait signature. Similarly, Tian et al. (2004) described the global and local features of shape contour using Fourier descriptors. In their work, three-dimensional Fourier transform was applied to the gait volume to obtain a unique frequency for individual walking pattern. Another variant was presented in Lu et al. (2008), where they represented every gait cycle as four key frames. Fourier transform was subsequently performed on these key frames to obtain the key frame profile. Elsewhere, Yuan et al. (2015) obtained five key frames from each gait cycle based on the ratios of silhouette bounding box. Fourier descriptors were thereafter deployed to describe the key frames silhouette boundary. Ohara et al. (2004) further introduced the idea of three-dimensional Fourier transform. Choudhury and Tjahjadi (2012) adopted both PMS and elliptical Fourier descriptors as gait signatures. The final label was decided by combining the outputs from both representations using rank-summation rule. In a more recent work, Lee et al. (2013) proposed optimally interpolated Fourier descriptors for gait recognition. Specifically, the closed contours of the body silhouette were circularly shifted by a circular permutation matrix before element-wise frame interpolation and Fourier transform was applied to produce length invariant gait signatures. Elsewhere, Boulgouris and Chi (2007) performed Radon transform on the silhouettes in each gait cycle. Subsequently, LDA was applied to identify the Radon coefficients that carry the most discriminative information. Table 4 summarizes transformation-based representation.

**Distribution-Based Representation**

In distribution-based representation, human walking is characterized by the statistical distribution generated throughout the gait cycle. Some of the more widely used representation in this category are optical flow distribution, probability distribution and texture-based distribution.

Polana and Nelson (1994) first adapted optical flow in the gait recognition problem. They tracked and recognized people walking in outdoor scenes by gathering the optical flow magnitudes and periodicity measurements over the entire body. Following that, Little and Boyd (1995) filtered the optical flow of human walking to produce a set of moving points and their flow values. Then, the geometry of the moving points was used to derive a gait signature. On the other hand, Bashir et al. (2009) computed the optical flow fields from preprocessed subject images over a gait cycle. Figure 14 depicts the computation of the optical flow field. In their representation, both the motion intensity and the motion direction information were captured in flow vectors. To achieve robustness against noise, the flow direction was discretised and a histogram-based direction representation was formulated. Lam et al. (2011), the optical flow field of the moving silhouettes was adopted to construct the Gait Flow Image (GFI) for gait recognition.

Apart from optical flow, probability distribution is also proposed as a gait representation. Vega and Sarkar (2003) modeled the relational statistics of gait image features in probability functions space, where each motion type creates a trace in this space. Recently, Hong et al. (2013) proposed a probabilistic gait representation. They considered the silhouette as a multivariate random variable and Bernoulli mixture model was employed to model silhouette distribution. Lee et al. (2014a) propounded yet another probabilistic gait representation by computing the binomial distribution of all pixels in the gait image. Thereafter, the mean and variance of the distribution is obtained.

![Fig. 14. The first two images show the silhouette at frame t – 1 and t respectively. The computed optical flow field is displayed in the third image (Bashir et al., 2009)](image-url)
Table 3. Summary of model-free approaches (appearance-based representation)

<table>
<thead>
<tr>
<th>Literature</th>
<th>Gait features</th>
<th>Classifier/distance metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobick and Davis (2001)</td>
<td>MEI + MHI</td>
<td>Mahalanobis distance</td>
</tr>
<tr>
<td>Lee et al. (2014b)</td>
<td>TAMHI + HOG</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Xu et al. (2006)</td>
<td>Averaged silhouette + CSA + DATER</td>
<td>kNN</td>
</tr>
<tr>
<td>Han and Bhanu (2006)</td>
<td>GEI</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Yang et al. (2008)</td>
<td>EGEI</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Huang et al. (2013)</td>
<td>Modified GEI + Gabor Wavelets</td>
<td>SVM</td>
</tr>
<tr>
<td>Zhang et al. (2009)</td>
<td>DGEI + PCA + Locality preserving projections</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Xu and Zhang (2010)</td>
<td>GEI + Fuzzy PCA</td>
<td>kNN + Euclidean distance</td>
</tr>
<tr>
<td>Moustakas et al. (2010)</td>
<td>GEI + Radial integration transform</td>
<td>Probability</td>
</tr>
</tbody>
</table>
| Xu et al. (2012)               | GEI + Gabor-PDF                                    | Locality constrained group sparse
| Zhang et al. (2010a)           | AEI + 2D locality preserving projections           | kNN + Euclidean distance                |
| Huang and Boulgouris (2012)    | SEI + Linear discriminant analysis                | -                                      |
| Chen et al. (2009)             | FDEI + Frieze + wavelet                            | HMMs                                   |
| Roy et al. (2012)              | PEI                                               | Euclidean distance                      |

Table 4. Summary of model-free approaches (transformation-based representation)

<table>
<thead>
<tr>
<th>Literature</th>
<th>Gait features</th>
<th>Classifier/distance metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murase and Sakai (1996)</td>
<td>Parametric eigenspace trajectories</td>
<td>Spatiotemporal correlation</td>
</tr>
<tr>
<td>Huang et al. (1999)</td>
<td>PMS + Canonical analysis</td>
<td>Spatiotemporal correlation</td>
</tr>
<tr>
<td>Wang et al. (2003a; 2003b)</td>
<td>PMS</td>
<td>Procrustes distance</td>
</tr>
<tr>
<td>Zhang et al. (2010b)</td>
<td>PMS + Shape context</td>
<td>kNN + Shape context distance</td>
</tr>
<tr>
<td>Zheng et al. (2011)</td>
<td>VTM</td>
<td>L1-norm distance</td>
</tr>
<tr>
<td>Kusakunniran et al. (2011a)</td>
<td>PSC</td>
<td>kNN + Procrustes distance</td>
</tr>
<tr>
<td>Kusakunniran et al. (2011b)</td>
<td>HSC</td>
<td>kNN + Procrustes distance</td>
</tr>
<tr>
<td>Mowbray and Nixon (2003)</td>
<td>Fourier descriptors</td>
<td>kNN + Euclidean distance</td>
</tr>
<tr>
<td>Tian et al. (2004)</td>
<td>Fourier descriptors</td>
<td>DTW</td>
</tr>
<tr>
<td>Lu et al. (2008)</td>
<td>Fourier descriptors (key frame profile)</td>
<td>kNN</td>
</tr>
<tr>
<td>Yuan et al. (2015)</td>
<td>Fourier descriptors (key frame)</td>
<td>Canonical Time Warping</td>
</tr>
<tr>
<td>Ohara et al. (2004)</td>
<td>3D Fourier descriptors</td>
<td>Cross correlation</td>
</tr>
<tr>
<td>Choudhury and Tjahjadi (2012)</td>
<td>PMS + elliptical Fourier descriptors</td>
<td>Procrustes distance + dissimilarity score</td>
</tr>
<tr>
<td>Lee et al. (2013)</td>
<td>Circular shifting + Interpolation + Fourier descriptor</td>
<td>Product of Fourier coefficients</td>
</tr>
<tr>
<td>Boulgouris and Chi (2007)</td>
<td>Radon transform + LDA</td>
<td>Euclidean distance</td>
</tr>
</tbody>
</table>

Table 5. Summary of model-free approaches (distribution-based representation)

<table>
<thead>
<tr>
<th>Literature</th>
<th>Gait features</th>
<th>Classifier/distance metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polana and Nelson (1994)</td>
<td>Optical flow</td>
<td>Nearest centroid</td>
</tr>
<tr>
<td>Little and Boyd (1995)</td>
<td>Optical flow</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Bashir et al. (2009)</td>
<td>Optical flow</td>
<td>Euclidean distance</td>
</tr>
<tr>
<td>Lam et al. (2011)</td>
<td>GFI</td>
<td>kNN + Euclidean distance</td>
</tr>
<tr>
<td>Vega and Sarkar (2003)</td>
<td>Trace in space of probability functions</td>
<td>Euclidean distance/DTW</td>
</tr>
<tr>
<td>Hong et al. (2013)</td>
<td>Multivariate probability + Bernoulli mixture model</td>
<td>Probability</td>
</tr>
<tr>
<td>Lee et al. (2014a)</td>
<td>Binomial distribution</td>
<td>Kulback-Leibler divergence</td>
</tr>
<tr>
<td>Kellokumpu et al. (2009)</td>
<td>LBO-TOP</td>
<td>Histogram intersection</td>
</tr>
<tr>
<td>Abdolahi and Gheissari (2011)</td>
<td>LBP-TOP + Histograms of video-words occurrences</td>
<td>SVM</td>
</tr>
<tr>
<td>Hu et al. (2013)</td>
<td>Optical flow + LBP</td>
<td>DTW</td>
</tr>
<tr>
<td>Lee et al. (2015)</td>
<td>TBP</td>
<td>Euclidean distance</td>
</tr>
</tbody>
</table>

The success of texture descriptors in face expression (Zhao and Pietikainen, 2007) and action recognition (Kellokumpu et al., 2008) inspired Kellokumpu et al. (2009) to engage texture descriptor based on Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) to represent human motion. They proposed a multi-resolution Local Binary Patterns (LBP) coding that is subsequently used to construct spatiotemporal LBP histograms. Abdolahi and Gheissari (2012) devised a rotation invariant version of the LBP-TOP. They extracted spatiotemporal interest points and described them by a dynamic texture descriptor. Afterwards, the gait signature was represented as a histogram of video-words occurrences. A hybrid of optical flow and texture...
Table 6. Summary of datasets used in vision-based gait recognition

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Segments</th>
<th>Subjects</th>
<th>Filming settings</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>USF Human ID gait baseline dataset</td>
<td>Large dataset</td>
<td>122</td>
<td>Outdoor</td>
<td>Viewpoint, Surface, Footwear,</td>
</tr>
<tr>
<td>Southampton Human ID at a distance gait dataset</td>
<td>Small dataset</td>
<td>12</td>
<td>Indoor</td>
<td>Footwear, Clothing, Carrying condition</td>
</tr>
<tr>
<td>CASIA Gait Dataset</td>
<td>Dataset A</td>
<td>20</td>
<td>Outdoor</td>
<td>Viewpoint</td>
</tr>
<tr>
<td></td>
<td>Dataset B</td>
<td>124</td>
<td>Indoor</td>
<td>Viewpoint, Clothing, Carrying condition</td>
</tr>
<tr>
<td></td>
<td>Dataset C</td>
<td>153</td>
<td>Outdoor, Infrared</td>
<td>Walking speeds, Carrying condition</td>
</tr>
<tr>
<td></td>
<td>Dataset D</td>
<td>88</td>
<td>Indoor + Footprint</td>
<td>-</td>
</tr>
<tr>
<td>CMU MoBo Gait Dataset</td>
<td>-</td>
<td>25</td>
<td>Treadmill</td>
<td>Viewpoint, Walking speeds, Carrying condition</td>
</tr>
<tr>
<td>OU-ISIR Gait Dataset</td>
<td>Treadmill dataset A</td>
<td>34</td>
<td>Treadmill</td>
<td>Carrying condition, Surface incline</td>
</tr>
<tr>
<td></td>
<td>Treadmill dataset B</td>
<td>68</td>
<td>Treadmill</td>
<td>Clothing</td>
</tr>
<tr>
<td></td>
<td>Treadmill dataset D</td>
<td>185</td>
<td>Treadmill</td>
<td>Gait fluctuations</td>
</tr>
<tr>
<td></td>
<td>Large population dataset</td>
<td>4016</td>
<td>Indoor</td>
<td>Viewpoint</td>
</tr>
</tbody>
</table>

Pattern Recognition and Classification

This section outlines some widely used pattern classification schemes in the gait recognition phase. The k-Nearest Neighbor (kNN) scheme (Fix and Hodges, 1951) is engaged when the gait features are encapsulated in a single representation, e.g., averaged silhouette. The kNN scheme is commonly based on the Euclidean distance between a test sample and the set of training samples. The predicted class of the test sample is set to the most frequent class label in the set of k nearest training samples with the minimum distance. Some examples of using kNN in gait recognition can be found in (Cunado et al., 2003; Wagg and Nixon, 2004; Wang et al., 2004; Xu and Zhang, 2010; Lu et al., 2008), among others.

Human walking is sometimes represented in a series of time-varying gait features, e.g., joint angle trajectories. The gait sequences are seldom realized at the same speed across the gait cycles, thus producing gait sequences of different lengths. A classification scheme that allows elastic shifting of the time axis (Keogh and Ratanamahatana, 2005) to minimize some distance measures is henceforth needed. To that end, the Dynamic Time Warping (DTW) technique (Berndt and Clifford, 1994) and its variants are usually applied to align gait sequences of different lengths (Tanawongsuwan and Bobick, 2001; Tian et al., 2004; Hu et al., 2013).

Inspired by the promising performance in speech recognition, state-space model such as the hidden Markov Model (HMM) was adapted for the gait recognition problem. HMMs are a widely adopted approach to the modeling of sequence data. The transitions between states and the generation of output symbols are determined by probability distributions (Stolcke and Omohundro, 1993). The application of HMMs in gait recognition can be found in Zhang et al. (2004; Chen et al., 2009).

Some researchers (Fathima and Banu, 2012; Abdolahi and Gheissari, 2012; Huang et al., 2013; Lu et al., 2014) applied multi-class Support Vector Machines (SVM) in the gait recognition problem. A multi-class pattern recognition problem is usually decomposed into multiple binary classification problems (Duan and Keerthi, 2005).

Neural network is likewise a widely used technique in pattern classification task. Zhang et al. (2005; Yoo et al., 2008; Xiao and Yang, 2008) employed back propagation neural network in gait recognition. Lee et al. (2008) applied an ensemble of neural network to achieve better generalization performance than a single neural network. Some other classification methods are also deployed. Fuzzy logic is frequently used in cases when there are overlapping characteristics. Given a test sample, the fuzzy logic method assigns proximity towards each training sample. The final label is determined based on the highest proximity value. The application of fuzzy logic in gait recognition could be found in Roy and Sural (2009; Bharti and Gupta, 2013), among others.

Gait Datasets

Standard gait datasets are required to evaluate the performance of gait recognition algorithms. In this section, several popular gait datasets, i.e., USF Human ID Gait Baseline Database, Southampton Human ID at a Distance Gait Dataset, CASIA Gait Dataset, CMU Motion of Body (MoBo) Dataset and OU-ISIR Gait Dataset, are briefly discussed. A summary of the datasets is presented in Table 6.
**USF Human ID Gait Baseline Dataset**

The USF Human ID Gait Baseline dataset (Sarkar et al., 2005) is collected at University of South Florida (USF). The dataset comprises 1870 video sequences of 122 people. For each person, there are 5 covariates, i.e., viewpoint, surface type, shoes, carrying condition and different time instants.

**Southampton Human ID at a Distance Gait Dataset**

There are two major segments in the Southampton Human ID at a distance gait dataset: Large dataset and small dataset. The large dataset has a population of more than 100 subjects with 3 scenarios: Indoor, outdoor and treadmill. The small dataset, on the other hand, contains video sequences of 12 subjects, filmed indoor with different covariates, i.e., footwear, clothing and carrying condition.

**CASIA Gait Dataset**

The Institute of Automation, Chinese Academy of Sciences (CASIA) publishes the CASIA gait dataset. There are four datasets in the CASIA gait database: Dataset A (standard dataset), Dataset B (multiview dataset), Dataset C (infrared dataset) and Dataset D (gait and footprint dataset).

The Institute of Automation, Chinese Academy of Sciences (CASIA) publishes the CASIA gait dataset. There are four datasets in the CASIA gait dataset: Dataset A (standard dataset), Dataset B (multiview dataset), Dataset C (infrared dataset) and Dataset D (gait and footprint dataset).

CASIA Dataset A (Wang et al., 2003a) consists of 20 subjects. Each subject has 12 image sequences with 4 sequences for each of the three directions, i.e., parallel, 45 degrees and 90 degrees to the image plane. CASIA Dataset B (Zheng et al., 2011; Yu et al., 2006), on the other hand, is a large multiview gait database consisting of 124 subjects and 11 views. Besides that, three variations are considered in the dataset, i.e., carrying condition, viewing angle and clothing. CASIA Dataset C (Tan et al., 2006) was acquired with an infrared camera at night. It contains 153 subjects with four walking conditions: Normal walking, slow walking, fast walking and normal walking with a bag. The last dataset, CASIA Dataset D, consists of 88 subjects. The dataset was collected synchronously by camera and Rscan Footscan.

**CMU MoBo Gait Dataset**

The CMU MoBo gait dataset (Gross and Shi, 2001) is collected by Carnegie Mellon University (CMU). The dataset consists of 25 individuals walking on a treadmill. Each subject performs four different walking patterns: Slow walk, fast walk, incline walk and walking with a ball. The dataset was captured by cameras placed at six different locations around the subject.

**OU-ISIR Gait Dataset**

The Institute of Scientific and Industrial Research (ISIR), Osaka University (OU) maintains two vision-based gait datasets, i.e., treadmill dataset and large population dataset.

The treadmill dataset comprises people walking on a treadmill surrounded by 25 cameras (Makihara et al., 2012). The treadmill dataset A contains gait sequences of 34 subjects with speed variation ranging from 2 to 7 km/h at 1 km/h interval. For each walking speed, it comprises 68 videos with two videos per subject. The treadmill dataset B contains gait sequences of 68 subjects with 32 different clothing combinations. The treadmill dataset C is still not publicly available yet. The treadmill dataset D consists of 370 gait sequences of 185 subjects. The dataset focuses on the gait fluctuations over time. The gait fluctuations were measured by Normalized Auto Correlation (NAC) of size-normalized silhouettes for the temporal axis. The dataset is divided into two subsets: DBhigh comprising 100 subjects with the highest NAC (stable gait) and DBlow comprising 100 subjects with the lowest NAC (fluctuating gait).

The large population dataset (Iwama et al., 2012) consists of 4016 people walking on the ground surrounded by 2 cameras at 30 fps, 640 by 480 pixels. Currently, only the dataset captured by the first camera (dataset C1) is publicly available.

**Concluding Remarks and Prospective Research**

This paper serves as a review of existing strategies in the feature extraction and pattern recognition stages of gait recognition. Previous work on feature extraction can be broadly grouped into two major categories, i.e., model-based and model-free. The model-based approaches explicitly model the human body by shapes and, thereafter, the properties of these shapes in a gait cycle are measured (structural model), or measures the motion via the kinematics of joint angles (motion model). Though model-based approaches are more robust to view and scale variations and reflect the kinematic characteristics of walking manner, they are difficult to accurately locate the joint positions due to the non-rigid structures of human body and to self-occlusion. Therefore, the current literature focuses more on model-free approaches.

The model-free approaches, on the other hand, directly operate on the gait sequences without assuming any specific model. The gait signatures in model-free approaches can be divided into appearance-based representation, transformation-based representation and distribution-based representation. The appearance-based representation captures the statistics (e.g., average, difference) of moving silhouette in the gait sequences. Some researchers reduce the dimension of the gait input
feature by projecting them onto other domains, giving rise to the transformation-based representation. The distribution-based representation, on the other hand, describes gait signatures by distribution or histograms.

Although much research has been devoted to gait recognition, the usability of gait recognition in a real application context still face challenges. Prospective research may consider to address several covariates, namely the viewing angles, clothing, carrying condition and speed. The majority of gait recognition algorithms are restricted to specific viewpoints and are sensitive to the change in viewing angles. Thus, combining different viewing angles as training data and transformation-based gait representation are some possible solutions. Clothing and carrying condition variations are of particular concern in real world applications. Some potential solutions are video acquisition using infrared camera, fusing gait with other biometrics such as face and transformation-based gait representation. Yet another challenge is the robustness to speed variations. Appearance-based and distribution-based gait representation might be appropriate to describe gait sequences of varying speeds.

Acknowledgment

Research reported in this paper was supported by Fundamental Research Grant Scheme, Grant No. MMUE/140073.

Author’s Contributions

Chin Poo Lee: Drafting the article.
Alan Wee Chiat Tan: Reviewing the article critically for significant intellectual content.
Kian Ming Lim: Reviewing the article critically for significant intellectual content.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References


DOI: 10.1109/AFGR.2004.1301502

DOI: 10.1109/TCSV.2003.821972

DOI: 10.1109/ICM.2003.1251144


DOI: 10.1109/ICNCSC.2008.4525252


DOI: 10.1109/TIP.2011.2160956


DOI: 10.1016/j.patcog.2003.09.012

DOI: 10.1016/j.sigpro.2008.03.006

DOI: 10.4218/etrij.11.1510.0068

DOI: 10.1109/IPTA.2008.4743792


DOI: 10.1109/ISCID.2015.135

DOI: 10.1109/ICMLC.2009.5212511

DOI: 10.1107/11539087_80

DOI: 10.1016/j.sigpro.2010.01.024

