MULTI-VIEW GAIT RECOGNITION WITH INCOMPLETE TRAINING DATA

Lan Wei[1]*, Yonghong Tian*[1*]*, Yaowei Wang*[2]*, Tiejun Huang*[1]

[1]School of EE & CS, Peking University, Beijing 100871, China
[2]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
*Corresponding to: yhtian@pku.edu.cn

ABSTRACT

Changes in the viewing angles pose a major challenge for gait recognition because the human gait silhouettes can be different under the various viewing angles. Recently, View Transformation Model (VTM) was proposed to tackle this problem by transforming gait features from across views to a common viewing angle. However, VTM must use the data of subjects crossing all views to train the pre-constructed model, which might be unsuitable for the real applications. To address this problem, this paper proposes a View Feature Recovering Model (VFRM) to generate the VTM with incomplete training data. In our algorithm, if the gait signature of a pedestrian is missing under a view, it can be recovered from the K-nearest pedestrians whose gait features are available in the same view. Moreover, the Geodesic distance based K-Nearest Neighbor (GKNN) algorithm is adopted in our algorithm to better measure the neighborhood between two pedestrians. Experimental results on a benchmark database has demonstrated the effectiveness of our method.

Index Terms— Gait recognition, View Transformation Model (VTM), View Feature Recovering Model (VFRM), Geodesic distance based K-Nearest Neighbor (GKNN), Incomplete data

1. INTRODUCTION

Human gait is one of the well-recognized biometric features for surveillance systems to ascertain the identity of a human at a distance from a camera. In practical visual surveillance scenarios, various factors can affect human gait, including variations of view, walking speed, carrying an object and even wearing different types of shoes [1]. Among them, changes in the viewing angles pose a major challenge for gait recognition. This is because the available visual gait features change a lot under different viewing angles. Moreover, the view under which the gait signatures database is generated may not be the same as the one when the probe data is obtained.

Basically, existing works on multi-view gait recognition can roughly be divided into three categories: 1) performing gait recognition under calibrated multi-camera system with view synthesis using 3D structure information; 2) extracting gait features that are invariant to view change; 3) projecting gait feature from one view to the other with view transformation. In the view synthesis based approaches, Shakhnarovich et al. [2] developed a view-normalization method for multi-view face and gait recognition. They utilized a set of monocular views to construct image-based visual hull (IBVH), from which to render virtual views for gait recognition. Zhao et al. [3] set up a human 3D model from video sequences captured by multiple cameras. They used the motion trajectories of lower limbs as dynamic features and exploited linear time normalization for matching and recognition. Ariyanto et al. [4] employed articulated cylinders with 3D Degrees of Freedom (DoF) at each joint to model the human lower legs. With gait structural and dynamics 3D features being extracted, a model-fitting algorithm was applied for gait recognition. Overall, these approaches often require a fully controlled and cooperative multi-camera system for reconstructing 3D gait model. Moreover, they are not applicable for real-time systems due to the high computation cost.

In the view-invariant features based approaches, Kale et al. [5] concluded that if a person was far enough from the camera, it was possible to generate a side view from any arbitrary view using a single camera. Thus they proposed a perspective projection model and flow-based structure from motion for gait recognition. Jean et al. [6] introduced an approach to compute and evaluate view-normalized trajectories of body parts from monocular video sequences. Extracting 2D trajectories of both feet and head as view-invariant gait features, the walking trajectory was segmented into piecewise linear segments. Goffredo et al. [7] utilized the angular measurements and trunk spatial displacement derived from the estimated lower limbs' poses as a view-invariant gait feature for viewpoint-independent gait biometrics. Kusakunniran et al. [8] proposed a method to normalize gaits from arbitrary views in the input layer (i.e., on gait silhouettes). The corresponding domain transformation obtained through invariant low-rank textures (TILTs) was adopted to transform sequences of gait silhouettes onto the common canonical view. Then, procrustes shape analysis (PSA) was applied on a sequence of the normalized gait silhouettes to extract a view-

This work is partially supported by grants from the National Natural Science Foundation of China under contract No. 61035001 and No. 61390515.

invariant gait feature with procrustes mean shape (PMS) and consecutively measure a gait similarity with procrustes distance (PD). Obviously, these methods work well only for a limited range of viewing angles. When two views differ a lot, their performance would be remarkably dropped.

In the view transformation based approaches, View Transformation Model (VTM) transformed gait features from different views into the same view. To construct the VTM, a training matrix was needed, in which each row contains gait features from the same view but different subjects, and each column contains gait features from the same subject but different views. Then, the matrix was decomposed into a view-independent matrix and a subject-independent matrix using singular value decomposition (SVD). The subject-independent matrix was used to construct the VTM. Makihara et al. [9] first introduced the VTM with gait features in the Frequency Domain. Then, Kusakunniran et al. [10] applied the truncated SVD (TSVD) to create the VTM and utilized Linear Discriminant Analysis (LDA) to optimize Gait Energy Image (GEI) as gait features. It had been proved that this method could improve the performance of the method in [9]. Later in [11], they re-formulated this problem as a regression problem. Support Vectort Regression (SVR) was proposed to create the VTM and local motion relationship was used to seek local Region of Interest (ROI) for predicting the corresponding motion information under different views. Zheng et al. [12] introduced a robust VTM via robust Principal Component Analysis (robust PCA). Gait feature was extracted by adopting the feature selection method with Partial Least Square (PLS) on the original GEI. As such, the roubst VTM was constructed while the view transformation projection and feature selection functions were learned. Overall, the view transformation based approaches can overcome the defects in the other two categories. However, they still have a limitation that they must use the data of pedestrians crossing all views to train the pre-constructed model, which might be unsuitable for the real applications.

To address this problem, we propose a View Feature Recovering Model (VFRM) for multi-view gait recognition with incomplete training data. In our method, if the gait signature of a pedestrian is missing under a view, it can be recovered from the neighborhood pedestrians whose gait features are available in the same view. This frees us from the need for gait features crossing all views in the processing of constructing VTM. Moreover, the Geodesic distance based K-Nearest Neighbor (GKNN) algorithm is adopted to obtain the better measured neighborhood between two pedestrians.

Experiments are conducted on the well-known benchmark database, the CASIA gait database B [13]. To evaluate the performance of the proposed VFRM, 10, 30 and 50 percent of the training data is randomly removed. The experimental results shows that with these incomplete data, our method can achieve the comparable or slightly lower performance with the VTM-based solution that is trained with complete data.

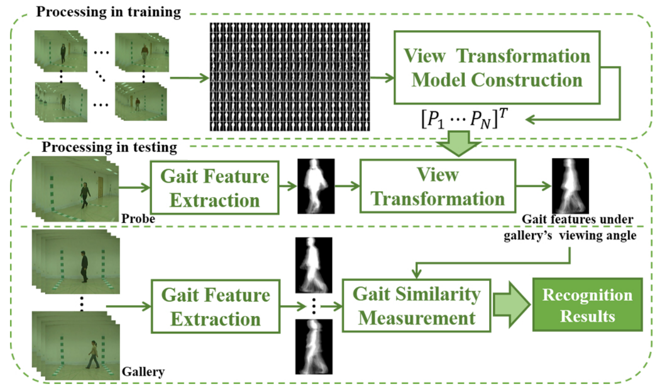

Fig. 1. Overview framework of VTM based Gait Recognition.

The rest of this paper is organized as follows. Section 2 describes the VTM-based solution, including the GEI extraction, VTM construction and gait similarity measurement. The proposed VFRM model and the GKNN algorithm are presented in Section 3. Experimental results are shown in Section 4 and the conclusion is drawn in Section 5.

2. VTM BASED GAIT RECOGNITION

The overview framework of VTM-based gait recognition is illustrated in Fig. 1. After gait features being extracted, the viewing angles of gallery gait data and probe gait data are transformed into the same direction with the generated VTM. Thus, gait signatures can be measured without difficulties.

2.1. Gait feature extraction

The first step for gait feature extraction is gait period estimation. We adopt the method in [10], which uses the the aspect ratio of silhouette bounding box to estimate the walking period. Consecutively, the well-known GEI [14], which has been reported as a good feature robust to silhouette errors and image noise, is extracted as the original gait feature. GEI is defined as:

$$g(x, y) = \frac{1}{T} \sum_{q=1}^{Q} \sum_{t=1}^{T} S_{q,t}(x, y) \tag{1}$$

where each $S_{n,t}(x, y)$ is a particular pixel located at position (x, y) of t_{th} $(t = 1, 2, \ldots T)$ walking silhouette image from q_{th} $(q = 1, 2, \ldots Q)$ gait cycle in a gait sequence. All silhouettes $S_{n,t}$ are rescaled along both horizontal direction and vertical direction to a fixed width (W) and height (H) respectively. T is the number of frames in gait cycle. S is a silhouette image at frame t. x and y are the image coordinates. The original extracted GEI feature representation is a 1-D vector, whose size is $W \times H$. Linear Discriminant Analysis (LDA) is applied on GEI to acquire an optimized version of the feature. The dimension of the gait features is reduced

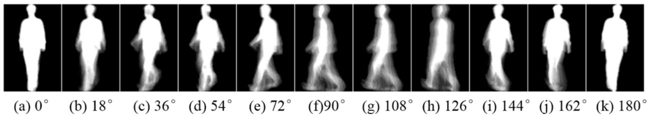

Fig. 2. GEIs of a same subject under different views.

to N_g. The extracted gait features for a subject in different viewing angles are shown in Fig. 2.

2.2. View transformation model construction

A gait matrix is created to construct VTM. Each row of the matrix contains gait features from same view but from the different subjects. Each column contains gait features from same subject but under the different views. There are total N viewing angles and M subjects for constructing VTM. g_k^m denotes the gait feature of subject m under n_{th} viewing angle. The factorization process by TSVD is as below:

$$\begin{bmatrix} g_1^1 & \cdots & g_1^M \\ \vdots & \ddots & \vdots \\ g_N^1 & \cdots & g_N^M \end{bmatrix} = USV^T = \begin{bmatrix} P_1 \\ \vdots \\ P_N \end{bmatrix} \begin{bmatrix} v^1 & \cdots & v^M \end{bmatrix} \tag{2}$$

where U is the $NN_g \times r$ orthogonal matrix. V is the $r \times r$ orthogonal matrix. S is the $r \times r$ diagonal matrix contains the first r $(r < m)$ largest singular values. $P = [P_1, ..., P_K]^T = US$ where P_k is the $N_g \times r$ sub-matrix of US. v^m is the r dimensional column vector. According to [10], the optimization of TSVD over regular SVD algorithm is that TSVD avoids over fitting problem by removing the less important elements from the transformation model.

The vector v^m is an intrinsic view-independent gait feature of the m_{th} subject. P_n is the subject-independent matrix which can project intrinsic gait feature vector v to the gait feature vector under specific viewing angle n. Thus, given the gait feature vector g_j^m from the m_{th} subject under the j_{th} viewing angle, the feature vector under the i_{th} viewing angle is easily obtained as:

$$g_i^m = P_i P_j^+ g_j^m \tag{3}$$

where P_j^+ is the pseudo inverse matrix of P_j.

2.3. Gait Similarity Measurement

With the generated VTM, gait features of gallery data and probe data from across views have been normalized into the same view. Thus, gait signatures can be measured using the simple L1-norm distance:

$$d(g_i, g_j) = \|g_i - g_j\| \tag{4}$$

where $d(g_i, g_j)$ is a distance between gait signatures g_i and g_j. The smaller value of $d(g_i, g_j)$ means the larger similarity between two gait signatures. The similarity of the two gait feature sequences is defined as:

$$D(Q_P, Q_G) = Median_i \left[\min_j \left\{ d(g_{P_i}, g_{G_j}) \right\} \right] \tag{5}$$

where $D(Q_P, Q_G)$ is the a distance between gait signature sequences in probe and gallery data.

3. VFRM BASED GAIT RECOGNITION WITH INCOMPLETE TRAINING DATA

In real environment, it is hard to obtain the gait features of a pedestrian crossing all viewing angles. In this section, we introduce the View Feature Recovering Model to recover the training matrix for VTM with incomplete training data.

3.1. Geodesic distances based K-nearest-neighbor

Assuming that when the gait feature of a subject has not been extracted under one view, at least one other subject's is available under the same view. The proposed VFRM recovered the missing data with the average of the k-nearest-neighbor subjects' gait features:

$$\begin{aligned} & g_t^i = \frac{1}{K} \sum_{k=1}^{K} g_t^k, \\ & g_t^k \in \left\{ g_t^k \,\middle|\, \begin{array}{l} g_t^k \neq 0, \quad k = 1, 2 \cdots K \\ D\left(M_i, M_1\right) \leq \cdots D(M_i, M_K) \end{array} \right\} \end{aligned} \tag{6}$$

where g_t^i denotes the missing gait feature of subject i under t_{th} viewing angle. $\left\{g_t^k\right\}$ represents the k-nearest subjects' gait feature under the same view. $D(M_i, M_j)$ is the distance between subject i and j, which will be addressed in the following subsection.

The rule under which the VFRM computes the average could also be replaced as "weighted average". That is,

$$\begin{aligned} & g_t^i = \sum_{k=1}^{K} w_{s_k} g_t^k, \\ & g_t^k \in \left\{ g_t^k \,\middle|\, \begin{array}{l} g_t^k \neq 0, \quad k = 1, 2 \cdots K \\ D\left(M_i, M_1\right) \leq \cdots D(M_i, M_K) \end{array} \right\} \end{aligned} \tag{7}$$

where w_{s_k} denotes the weight of the k_{th} subject's gait feature. w_{s_k} is defined as:

$$w_{s_k} = \frac{1/D(M_i, M_k)}{\sum_{j=1}^{K} 1/D(M_i, M_j)} \tag{8}$$

In this paper, we just apply the normal average in Formula. 6.

3.2. Pedestrians neighborhood measurement

The key problem is how to measure the neighborhood between two subjects. Firstly, we utilize Dijkstra's algorithm to

1. Compute the Euclidean distance matrix.
Put all features available in a vector:
$G = \left\{ \left\{ g_x^i \left| g_x^i \neq 0, i \in V \right. \right\}, \left\{ g_y^j \left| g_y^j \neq 0, j \in V \right. \right\} \right\}$.
Assume the size of vector is N_G. Compute the Euclidean distance matrix between any two features in G;
for $r_1 = 1; r_1 < N_G; r_1 + +$ **do**
  for $r_2 = r_1 + 1; r_2 < N_G; r_2 + +$ **do**
    $d_e(r_1, r_2) = \sqrt{(G(r_1) - G(r_2))^2}$;
  end
end
2. Compute neighborhood graph A.
Initialize $N = 0$;
for $r_1 = 1; r_1 < N_G; r_1 + +$ **do**
  for $r_2 = r_1 + 1; r_2 < N_G; r_2 + +$ **do**
    if r_2 *is one of the* K_g*-nearest points of* r_1 **then**
      set $A(r_1, r_2)$ connected;
      $d_g(r_1, r_2) = d_e(r_1, r_2)$;
    end
    else
      $d_g(r_1, r_2) = \infty$;
      $N + +$;
    end
  end
end
3. Approximate geodesic distances with the shortest path in graph A.
for $n = 0; n < N; n + +$ **do**
  $d_g(r_1, r_2) =$
  $\min\{d_g(r_1, r_2), d_g(r_1, k) + d_g(k, r_2)\}$;
end
Finally $D_g = \{d_g(r_1, r_2)\}$ will contain the shortest path distances between all pairs of points in A, which outputs the desired geodesic distances matrix.

Algorithm 1: Geodesic distances computing algorithm

approximate the geodesic distances between any two gait features from different subjects. The reason why we use geodesic distances rather than Euclidean distances is that, when the two gait features are extracted from different views, it's more reasonable to consider the data distributed in a high dimensional manifold space. The geodesic distances measurement method is similar to steps in [15], which is described in Algorithm 1.

After achieving the geodesic distances between any two gait features from different subjects, the desired distance of subjects is extracted by weighted averaging the distances of features:

$$D(M_i, M_j) = \sum_{n_i=1}^{V} \sum_{n_j=1}^{V} w_{n_i n_j} d(g_{n_i}^i, g_{n_j}^j), g_{n_i}^i \neq 0, g_{n_j}^j \neq 0 \quad (9)$$

We also introduce a novel method to assign the weights

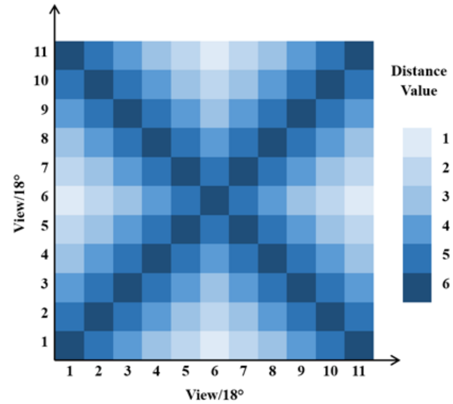

Fig. 3. Distance between any two viewing angles.

of the distances between the features. As illustrated in Fig. 2, the gait features from the same view contains same gait information, as well as the mirror view (e.g. the mirror view of 54°is view 126°). The more two views differ, the less common their features have in. So, with rules: 1) the distance of gait features from the same view gets heavier weight, as well as the mirror view; 2) the bigger difference between two viewing angles, or between one and the other's mirror angle, the lighter weight is assigned to the distance:

$$dw_{n_i n_j} = \frac{N+1}{2} - \min\{|n_i - n_j|, |N - n_i + 1 - n_j|\} \quad (10)$$

where $dw_{n_i n_j}$ represents the "distance" between two views. N is the number of viewing angles. The distances are drawn in Fig. 3, in which the deeper blue means a bigger value of $dw_{n_i n_j}$. The weight of distances between gait features is defined as:

$$w_{n_i n_j} = \frac{dw_{n_i n_j}}{\sum_{n_i=1}^{N} \sum_{n_j=1}^{N} dw_{n_i n_j}} \quad (11)$$

4. EXPERIMENTS

Dataset– The CASIA gait database B [13] is used in our experiments. The dataset consists 124 subjects from 11 viewing angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°). There are 10 walking sequences consisting of 6 normal walking sequences, 2 carrying-bag sequences and 2 wearing-coat sequences for each subject from each view. 6 walking sequences from 9 views from 18°to 162°are used in our experiments. Because the approximate frontal views 0°and 180°provide gait information which is too different from the canonical view. 24 subjects in the dataset are used to train VTM. The rest of 100 subjects are used for multi-view gait recognition. Specifically, the 6 walking sequences of each subject are separated into probe and gallery group, which contains 4 and the last 2 walking sequences respectively. Gait recognition rate is employed to evaluate the correct matching numbers gait feature vectors in the gallery dataset with probe data.

Multi-View Gait Recognition with Incomplete Training Data– To evaluate the performance of the proposed VFRM, we randomly remove 10, 30 and 50 percent of the data for training VTM in three experiments, respectively. VFRM introduced in Section 3 is subsequently applied to recover the data. The recovered matrix with 70% training data is drawn in Fig. 4, which contains unambiguous gait features from different subjects crossing all views. Fig. 5 illustrates the first rank multi-view gait recognition by using gait recognition method described in Section 2. Each bar chart is computed by transforming probe gait data to a feature set under viewing angle that matches one of the views in gallery gait data. Then, L1-norm distance is employed to measure the similarity. It can be seen from Fig. 5 that, with 10% or 30% data missing, our method achieve the comparable performance with the VTM-based solution that is trained with complete data. Occasionally, the proposed methods even outperforms the recognition with complete data. This is because the VFRM replaced the imperfect gait features, which are extracted from a set of fragmentary walking silhouettes, with better features containing more gait information. While, when 50% data is unavailable, the raise of redundant information resulting in slightly lower performance with the proposed method.

VFRM with Different Distance Measurements– VFRM using Euclidean distance based KNN rather than GKNN is also tested in our experiments. The recovering error rate is defined as:

$$r = \frac{\sum_{v=1}^{V} \sum_{m=1}^{M} \|g'^{m}_{v} - g^{m}_{v}\|}{\sum_{v=1}^{V} \sum_{m=1}^{M} \|g^{m}_{v}\|} \tag{12}$$

where g'^{m}_{v} and g^{m}_{v} are the recovered gait feature and original gait feature respectively. r is the recovering error rate with the corresponding distance measurement method. The smaller value of r means the larger similarity between the recovered data with the original data, and the better performance of the corresponding method. Results of two different distance measurement methods are shown in Table 1. As exhibited in Table 1, the proposed GKNN approach outperforms than KNN. Since the geodesic distances helps to provide a more accurate description about the distance between gait features from different viewing angles, it is not surprise that it performs better. While, results of the two methods haven't made a big difference when a small percent of data is missing. Besides, the Euclidean distance based KNN has advantages including fast processing and small data storage than GKNN. Thus, the Euclidean distance based KNN is also an alternative choice for VFRM when there is not too much data missing.

Table 1. Recovering error rate with different methods

Data Percent	10	30	50
GKNN	0.026	0.074	0.098
KNN	0.025	0.077	0.111

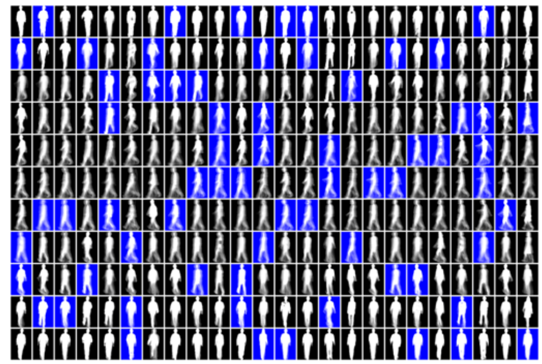

Fig. 4. Gait feature matrix recovered by VFRM with 70% training data. Subgraphs with black background are the available GEIs; ones with blue background are the recovered GEIs.

5. CONCLUSIONS

In this paper, we present a novel View Feature Recovering Model for multi-view gait recognition with incomplete training data. VFRM outputs the matrix containing gait features of subjects crossing all views to construct VTM. With VTM transforming gait signatures under various views into a common view, the similarity is measured easily. The proposed approach has been testified on a large multiple views gait database. It is shown to be an efficient method for multi-view gait recognition with incomplete training data. In the future, we plan to verify the proposed method in a more difficult dataset such as the HumanID Gait Challenge Problem Datasets [1] and seek a new data recovery scheme to further improve the performance.

6. REFERENCES

[1] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. PAMI.*, vol. 27, no. 2, pp. 162–177, 2005.

[2] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *CVPR*, 2001, vol. 1, pp. I–439.

[3] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3d gait recognition using multiple cameras," in *AFGR*. IEEE, 2006, pp. 529–534.

Fig. 5. Results of multi-view gait recognition performances with the proposed method between incomplete training data .

[4] G. Ariyanto and M. S. Nixon, "Model-based 3d gait biometrics," in *Biometrics (IJCB)*. IEEE, 2011, pp. 1–7.

[5] A. Kale, A. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Conf. AVSS*, 2003, pp. 143–150.

[6] F. Jean, R. Bergevin, and A. B. Albu, "Computing and e-valuating view-normalized body part trajectories," *IVC*, vol. 27, no. 9, pp. 1272–1284, 2009.

[7] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-calibrating view-invariant gait biometric-s," *IEEE Trans. on SMC- part B*, vol. 40, no. 4, pp. 997–1008, 2010.

[8] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, and H. Li, "A new view-invariant feature for cross-view gait recognition," 2013.

[9] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *ECCV*, pp. 151–163. 2006.

[10] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in *ICCV Workshops*, 2009, pp. 1058–1064.

[11] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Support vector regression for multi-view gait recognition based on local motion feature selection," in *Proc. IEEE Conf. CVPR*, 2010, pp. 974–981.

[12] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *Proc. IEEE Conf. ICIP*, 2011, pp. 2073–2076.

[13] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *ICPR*, 2006, vol. 4, pp. 441–444.

[14] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Conf. CVPR*, 2004, vol. 2, pp. II–842.

[15] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.