

Gait Style and Gait Content: Bilinear Models for Gait Recognition Using Gait Re-sampling

Chan-Su Lee
Department of Computer Science
Rutgers University
New Brunswick, NJ, USA
chansu@caip.rutgers.edu

Ahmed Elgammal
Department of Computer Science
Rutgers University
New Brunswick, NJ, USA
elgammal@cs.rutgers.edu

Abstract

Human Identification using gait is a challenging computer vision task due to the dynamic motion of gait and the existence of various sources of variations such as viewpoint, walking surface, clothing, etc. In this paper we propose a gait recognition algorithm based on bilinear decomposition of gait data into time-invariant gait-style and time-dependent gait-content factors. We developed a generative model by embedding gait sequences into a unit circle and learning nonlinear mapping which facilitates synthesis of temporally-aligned gait sequences. Given such synthesized gait data, bilinear model is used to separate invariant gait style which is used for recognition. We also show that the recognition can be generalized to new situations by adapting the gait-content factor to the new condition and therefore obtain corrected gait-styles for recognition.

1 Introduction

Human gait is a valuable biometric cue that can be used for human identification besides other biometrics such as faces and fingerprints. Gait has advantages compared to other biometrics [2] as it is easily observable in an unintrusive way and is difficult to disguise. Therefore, gait recognition has become more attractive for surveillance and for security in public areas [2, 8, 15, 7, 1]. However, there are several problems to be solved in order to build an automated vision-based gait recognition system that is useful in real applications. Most challenging is the existence of various sources of variations that might affect the appearance of gait in image sequences such as viewpoint, clothing, walking surface, shoe type, carried objects, etc.

The appearance of gait in image sequences is a spatiotemporal process that characterizes the walker. Gait recognition algorithms, generally, aim to capture discriminative spatiotemporal features from image sequences in or-

der to achieve human identification. We can categorize gait-recognition approaches into model-based approaches and appearance-based approaches according to the features they use for classification. Model-based approaches [2, 10, 8] fit models or intermediate body representations in order to extract proper features (parameters) that describes the dynamics of the gait. Appearance-based approaches aim to capture a spatiotemporal gait characteristics directly from input sequences. Murase [14] used parametric eigenspace representation to represent moving object using Principle Component Analysis (PCA). Huang *et al.* [7] extended the method using Canonical space transformation (CST) based on Canonical Analysis (CA), with eigenspace transformation for feature extraction. BenAbdelkader *et al.* [1] used self similarity measures to capture spatiotemporal characteristics using PCA analysis. Hayfron-Acquah *et al.* [5] used symmetric information to capture gait motion. Little *et al.* [11] computed phase vector from extracted optical flow. Shutler *et al.* [18] used higher order moments. In [6] HMM was used to capture gait dynamics from quantized Hu moments of silhouettes. HMM was also used in [9] with features representing silhouette width distribution. Still, it is difficult to extract good features to capture gait characteristics of individual people.

In this paper we introduce a new approach for gait recognition that follows the appearance-based paradigm. It is well known in psychology that human perceptual systems naturally separate the content and style factors of their observation in identifying a familiar face or gait seen under unfamiliar viewing conditions. The pioneering work of Tenenbaum and Freedman [20] applied a bilinear model to discover explicit parameterized representation for separable two-factor models such as different people faces and different head poses.

In the context of gait we aim to separate two orthogonal factors: *gait style*: time-invariant personalized style of the gait which can be used for identification; and *Gait*

content: time-dependent factor representing different body poses during the gait cycle. In order to achieve such separation, we developed a generative model to capture gait dynamics and represent gait image sequence invariant to walking speed. Gait image sequences can be thought as sampling from continuous gait motion. By learning a generative model we can reconstruct the image sequence at any time instances, and therefore we can synthesize intermediate gait poses. We use standardized embedding of gait image sequences into two-dimensional unit circles which is inspired by the embedding of gait manifold to low dimensional spaces using nonlinear dimensionality reduction frameworks like locally linear embedding (LLE) [17]. We generate new gait image sequence from this low dimensional embedding by learning nonlinear mapping using generalized radial basis function (GRBF) [16]. Generated new image sequences are temporally-aligned and, therefore, facilitate fitting of a bilinear model to separate the gait style and gait content. Gait style factor is used as feature vector in human identification using support vector machine (SVM) and other classifiers. The classification result shows good performance using gait style vectors. We also show that the recognition can be generalized to new situations, such as new view point, by adapting the gait-content factor to the new condition and therefore obtain corrected gait-styles for recognition.

We use NIST-USF Gait database [15] to test our approach. We use computed silhouettes for the May-Nov-2001 data. This data set has probe sets with variation of viewpoint, footwear, walking surface and with/without briefcase. We tested for the variation of viewpoint, footwear, and walking surface and compared with baseline result [15]. Section 2 shows how synthesized, temporally aligned gait data can be obtained. Section 3 shows how to learn two separate factors from gait image sequences. Section 4 describes the recognition process and adaptation to new situations. Gait recognition result will be shown in Section 5 prior to conclusion in Section 6.

2 Synthesized Gait

This section describes how to generate temporally aligned gait pose images for use in recognition using a generative model learned from the input sequences. Let us consider human silhouette during a gait cycle. Obviously, the human gait evolves along a one dimensional manifold embedded in a high dimensional visual space. Such manifold is nonlinear and can be twisted and self intersecting on the high dimensional space given the view point, person shape, clothing, etc. We studied the dimensionality and the geometric structure of the gait manifold in [3, 4]. We applied nonlinear dimensionality reduction frameworks such as locally linear embedding (LLE) [17] and isometric feature mapping (Isomap) [19] to gait image sequences obtained

from different views to discover the geometric structure of the gait manifold as well as to establish low dimensional embedding of such manifold. The gait manifold, generally, can be embedded in a three-dimensional Euclidean space discriminatively (i.e., we can discriminate different body poses during the gait cycle as laying at different points in the space) [3]. Only one degree of freedom controls the walking cycle which corresponds to the constrained body pose as a function of time. Unfortunately the twisting of the manifold depends on the view point, person body built, clothing, etc. On the other hand, half gait cycles can be embedded in a two dimensional Euclidean space as ellipse-like curves. Figure 1 shows example of such embedding. The first row in the figure 1 shows embedded gait manifold of side views for four different people. The second and third rows show embedding of half cycles.

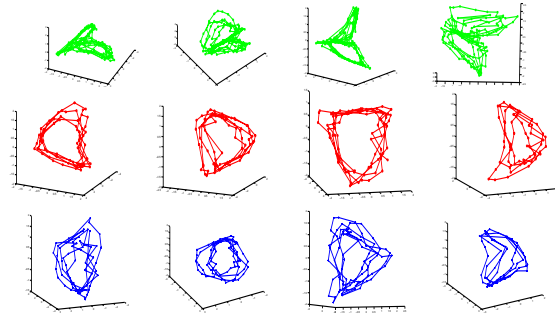


Figure 1. Low dimensional embedding for gait cycles by LLE

In order to achieve a recognition task with the existence of such twists in the embedded manifolds, we need to use a standardized embedding that approximates the manifold embedding. Therefore, we embed each half gait cycle temporally on a unit circle, i.e. a one dimensional manifold embedded in a two dimensional space. Input silhouettes corresponding to each half walking cycle are embedded on an equally spaced points along a unit circle.

Given such embedding of a half walking cycle, we need to synthesize new silhouettes at standard time instances during the cycle to be used for recognition. We define *N-synthesized gait poses* as a collection of *N* synthesized silhouettes at *N* equally spaced time instances during half a cycle which indicates how the silhouette shape will look like at these *N* standard intermediate points.

In order to obtain such synthesized gait poses, we learn a nonlinear mapping function from the manifold embedded on a unit circle into the input silhouettes. Learning nonlinear mapping is necessary since the manifold is embedded nonlinearly and arbitrary into a unit circle. We use generalized radial basis function (GRBF) [16] to learn such mapping as a collection of interpolation functions.

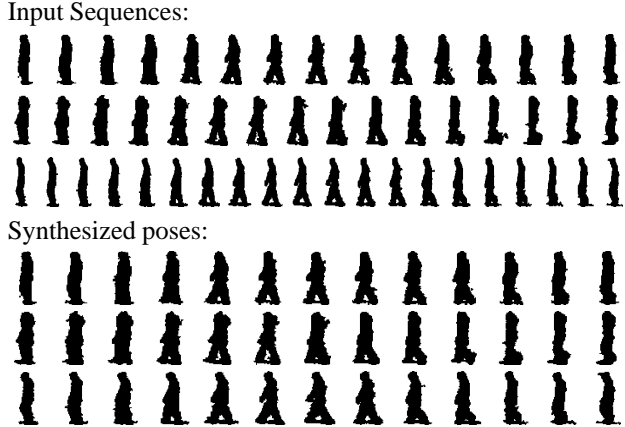


Figure 2. Original gait image sequences and their normalized gait poses

Let N equally spaced centers along a unit circle be $\{t_j \in R^e, j = 1, \dots, N\}$ and given a set input images $Y = \{y_i, i = 1, \dots, M\}$ and let their corresponding embedding along the unit circle be $X = \{x_i, i = 1, \dots, M\}$, we can learn interpolants in the form

$$f^k(x) = p^k(x) + \sum_{i=1}^N w_i^k \phi(|x - t_i|), \quad (1)$$

that satisfies the interpolation condition $y_i^k = f^k(x_i)$ where y_i^k is the k -th pixel of input silhouette y_i , $\phi(\cdot)$ is a real valued basic function, w_i^k are real coefficients, $p^k(\cdot)$ is a linear polynomial, and $|\cdot|$ is the norm on R^2 . The mapping coefficients can be obtained by solving a linear system of equations as shown in [3]. Such mapping can be written in the form of a generative model as

$$f(x) = B \cdot \psi(x) \quad (2)$$

that nonlinearly maps any point x from the two dimensional embedding space into the input space and therefore can be used to synthesize N intermediate silhouettes at N standard time instances equally spaced along the unit circle. In figure 2, first three rows show original image sequences for three different people and next three show 13-synthesized gait poses synthesized using the learned models from each input sequence.

In order to show the effect of synthesized gait poses in recognition we used a naive gait recognition algorithm using the synthesized gait data to evaluate the effectiveness of gait cycle normalization in recognition. In this experiment, we used 13-synthesized gait poses representing one half cycle for each person as feature vector for gait classification. We used eight half cycles to represent each person. Given a probe sequence, half cycles are automatically detected, embedded, and a model is learned in the form of equation 2. 13-synthesized gait poses for each probe half cycle are used

Difference	Probe Set	Baseline	N-Synthesized Gait
View	GAL	73%	100%
Shoe	GBR	78%	79%
Shoe, view	GBL	48%	50%
Surface	CAR	32%	57%
surface, view	CAL	17%	36%
Surface, shoe	CBR	22%	57%
Surface, shoe, view	CBL	17%	36%

to represent the probe subject. Classification is based on comparing 13-synthesized gait poses as a feature vector using simple Euclidean distance and nearest neighbor classification. Table 2 shows the human identification results for 14 people from NIST-USF Gait database. The table simply shows that classification based on naive image comparison using the synthesized gait poses can result in improvements over the basic baseline algorithm on the same data.

3 Bilinear Model for Gait

It is well known in psychology that human perceptual systems naturally separate the content and style factor of their observation in identifying a familiar face or gait seen under unfamiliar viewing conditions. The pioneering work of [20] showed that such separation can be achieved using a bilinear model with application to face images. Bilinear models are introduced as an extension of two-mode component analysis for scalar element in psychometrics [12] to vector elements using singular value decomposition (SVD) in [13] [20].

In the context of gait we aim to separate two orthogonal factors: *gait style*: time-invariant personalized style of the gait which can be used for identification; and *Gait content*: time-dependent factor representing different body poses during the gait cycle. Gait content is also dependent on other conditions such as viewpoint, shoe, ground, etc. An input silhouette can be represented by a bilinear model as

$$I_{sc} = \sum_{i=1}^N \sum_{j=1}^J w_{ij} c_i s_j \quad (3)$$

using gait style vectors s and gait content vectors c and basis images w_{ij} , i.e., the model linearly combines basis images w_{ij} using the style coefficients s_j and content coefficient c_i . The gait content vector varies with time through the walking cycle to generate the various body poses observed through the walking given the time-invariant gait style vector that characterizes the walker.

Given a training data with different people and multiple gait cycles per person which might manifest different conditions, the objective is to fit a model in the form of equation 3. The first step towards this is to warp the time domain of different cycles to establish correspondences in time between different cycles. This is done by embedding each cycle on a unit circle as shown in section 2 and therefore

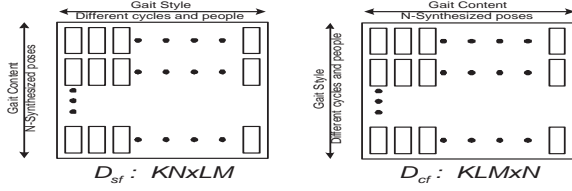


Figure 3. Style format and Content format

synthesize intermediate poses at standard time instances to represent each cycle in the training set.

Given L Gait cycles for each of M different people in the training data where each gait cycle is represented as N synthesized time-aligned poses at N standardized time instances, where each image is represented as a K dimensional vector, we aim to fit a symmetric bilinear model in the form of equation 3.

We arrange the synthesized gait image sequence into two forms: one is *style format*, D_{sf} , the other is *content format*, D_{cf} , as shown in figure 3. In style format, we have LM columns where each column contains N synthesized gait pose images as one gait cycle vector and the column vector size is KN . In content format, we have N columns where each column represents images of the same synthesized gait pose from all of the different gait cycles and different people gait sequences, i.e., each column is of KLM dimension.

Given such arrangement, the objective is to decompose the style and content vectors, i.e., to decompose the matrix D_{sf} as

$$D_{sf} = CW_{cs}S \quad (4)$$

or similarly $D_{cf} = SW_{sc}C$. Such model is called a symmetric bilinear model and it is necessary in order to adapt the gait styles to new gait contents given new situations as will be discussed later. In order to achieve such decomposition, Asymmetric bilinear model is used to decompose the data to separate gait style vectors S given content-dependent mapping T_c and to separate gait content vectors C given style-dependent mapping T_s as

$$D_{sf} = T_c S \quad (5)$$

$$D_{cf} = T_s C \quad (6)$$

to minimize the reconstruction error, i.e., to minimize $E = \|D_{sf} - T_c S\|^2$ and similarly for D_{cf} . Such decomposition can be achieved by singular value decomposition (SVD) as was shown in [20]. Given SVD for D_{sf} as $D_{sf} = UDV^T$, least square optimal solution is $S = V^T$ and $T_c = UD$. Similarly we can achieve the decomposition in equation 6. We can use J -dimensional approximation by choosing first J largest diagonal terms in D and setting the rest to zeros.

Given an asymmetric model fitted in the form of equation 5 and 6, symmetric model can be fitted iteratively.

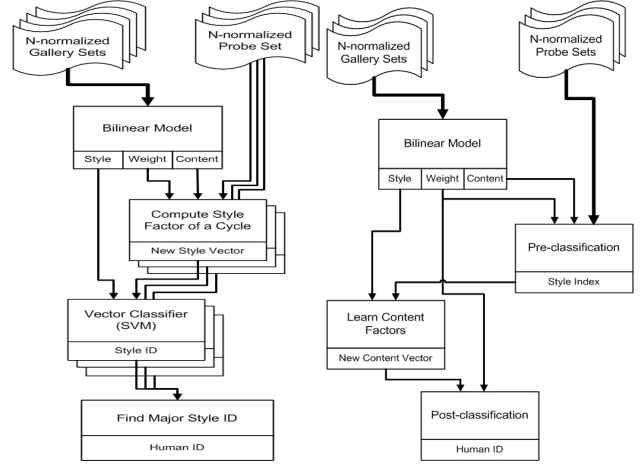


Figure 4. Recognition Algorithm

4 Gait Recognition Algorithm

Given a new probe sequence and given a learned model from the training data, the objective is to identify the person in the probe sequence, i.e., to recognize the gait style. Each probe sequence is first segmented into half cycles and each half cycle is embedded into a unit circle as was described in section 2. Then, nonlinear mapping is used to synthesize time-aligned poses to be used in recognition. Given N -synthesized silhouettes for each cycle of the probe sequence, the data is arranged into a column vectors I_{probe}^k of dimensionality KN where K is image size (height \times width) for each probe cycle k . Given the asymmetric model learned from the training data as in equation 5, we can solve for style vectors s_{probe}^k using the pseudo-inverse for the content-dependent mapping T_c , i.e.,

$$s_{probe}^k = T_c^{-1} I_{probe}^k = D^{-1} U^T I_{probe}^k$$

The resulting probe style vectors, and the style vectors learned from the training data are the basic features that can be used in the recognition. Each gait style vector is a point in a J -dimensional feature space, and general classifiers can be used for classification purpose. We used both k-NN classifier and SVM classifier to classify any new probe-style vector to one of the learned people classes. The recognition procedure is shown in figure 4.

Notice that such recognition procedure uses only the asymmetric model in equation 5. Why then we need a symmetric model? The answer is that we need symmetric model to adapt the model to new environment and situations as will be discussed next.

4.1 Adapting Gait Style to New Situations

We expect gait style factor to be invariant to different situations such as view point, shoe, ground, etc. How can that be achieved if we do not see all these different situations in the training data? Given a learned model using

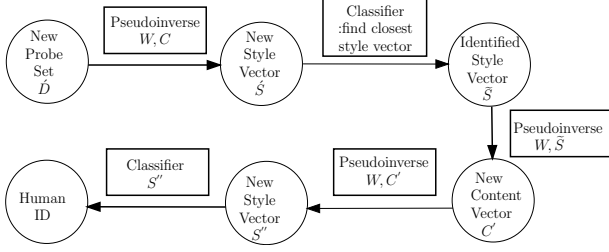


Figure 5. Classifier with adaptation of content vectors

data collected under certain situations, how can we use such model for recognition under different situations? new situation means new gait content or simply means new content-dependent mapping. Given a symmetric model in the form of equation 4, if we can adapt the content vector C to a new content vector C' for the new situation we can then solve for the style vectors under the new situation. In other words we need to extrapolate gait styles given new situations.

Given a probe data set captured under new condition, we show here how to recognize the people in the probe set by adapting the model to the new condition. The only assumption we make is that all the people in the probe set are part of the gallery set used for the training (This assumption is necessary anyway to be able identify the people in the probe set). If we know the correspondences between people in the probe set and the people in the training set, then, we can obviously solve for the new content vectors C' using the inverse of the style-dependent mapping T_s . Unfortunately, we do not know such correspondences since we do not know people class for the probe set (this is the recognition problem we want to solve!).

In order to simultaneously identify people and adapt to new situations we use the following procedure. First, the probe data set is arranged in style-format and content-format, $D_{sf}^{probe}, D_{cf}^{probe}$ as was shown in section 3 after detecting cycles, embedding, and gait synthesis as was shown in section 2. We can solve for approximate style vectors S' for the probe set by pseudo-inverse using original W, C as

$$S' = [(W^\nu C)^\nu]^{-1} D_{sf}^{probe} \quad (7)$$

where ν is matrix vector-transpose operation as defined in [20]. Given the recovered styles S' we can classify each cycle in the probe set to identify corresponding person from the training data. We call this step pre-classification. Given the pre-classification result we can recover the original style vectors \tilde{S} by finding closest style vector in the original style vectors for each probe cycle. Now, we can adapt content vectors to a new situation by solving for the new content vectors C' using the recovered style vector as

$$C' = [(W\tilde{S})^\nu]^{-1} D_{cf}^{probe} \quad (8)$$

This adapted content vectors C' are expected to represent

new environment better than original content vectors C . Finally, we can use the adapted content vectors C' to obtain new style vectors S'' in the same way as in equation 7 which is then can be used for final classification in the new environment. The procedure is shown in figure 5. This iterative procedure can be repeated to obtain better results.

5 Experimental Result

NIST-USF Gait database [15] is used to learn and evaluate our algorithm. We used computed silhouettes for the May-Nov-2001 data. We arbitrary select 14 peoples for preliminary evaluation from grass surface, shoes type A, and right camera sequences as a gallery set (GAR) and tested seven different conditions by variation of viewpoint (L), footwear (B), and walking surface (C). Original image size is 128×88 and were resized 64×44 , i.e., each input vector size is $64 \times 44 = 2816$. The number of gait poses in synthesized gait is 13 and vector size for one gait cycle is $36608 = 2816 \times 13$. The number of style vectors are $112 = 8(\text{cycle per person}) \times 14 (\text{people})$. The dimension of the data set D_{sf} is 36608×112 .

We evaluated identification accuracy with four different approaches: 1) An asymmetric model with nearest neighbor classifier on the recovered gait styles. 2) A Symmetric model with nearest neighbor classifier. 3) A Symmetric model with k-nearest neighbor classifier (k-NN). 4) A Symmetric model with support vector machine classifier (SVM). For identification using gait, we need to determine people based on sequences which might be composed of several cycles. So the classifiers we used identify people from each available input cycle, and boost the result of multiple cycles by selecting the majority of individual classification results. Figures 6-a,b show classification rates using the different classifiers.

The adaptation of the gait style to new environment helps classification in variant situation. Figures 6-c,d show comparison of human identification accuracy of a symmetric model without adaptation of gait content vector and with adaptation of gait content vector using nearest neighbor as well as using SVM classifier. In most of the cases, improvement can be noticed using adaptation of content vector to new situation. In the cases where pre-classification results are above 50%, the final gait recognition results show improvement because good pre-classifications make it possible to estimate the original style vectors well and, therefore, the new situation content vector can be recovered which leads to improvements in the final style classification results.

6 Conclusion

We presented new approach to gait recognition problem using low dimensional embedding and bilinear model

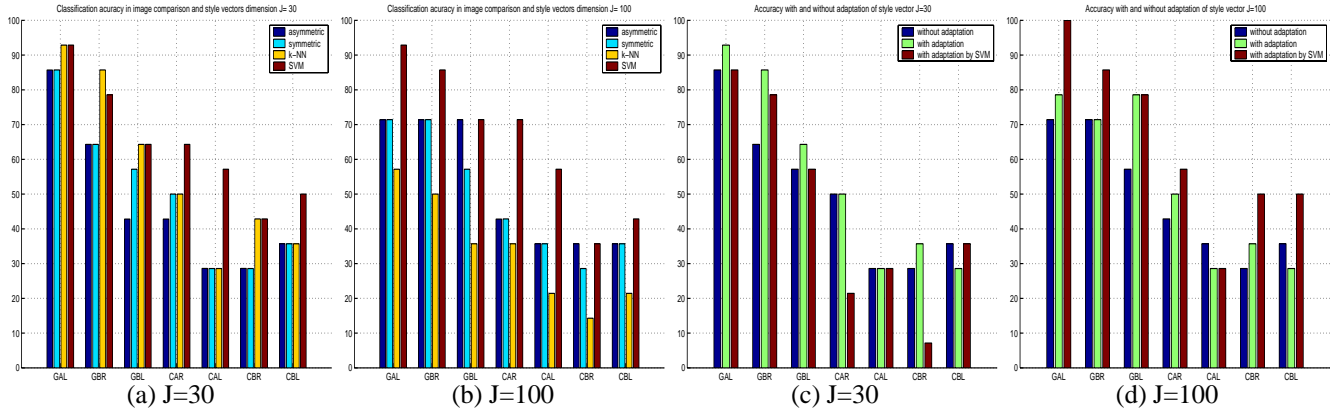


Figure 6. a,b: classification rates based on recovered style vectors. c,d: classification rates with/without adaptation to new content vectors

fitting. Low dimensional embedding of gait image sequences to unit circles enables us to generate time-aligned gait pose images, which are sampling invariant. Using bilinear model, we get representation of gait style which is time-invariant as well as invariant to different situations in the training data. Bilinear model facilitates generalization to new situations by adaptation of gait content vector to new probe set. Using gait style vectors and SVM classifier, we show promising human identification compared to the baseline algorithm. In the future we plan to report more large scale evaluations on more gait data sets.

Acknowledgment This research is partially funded by NSF award IIS-0328991

References

- [1] C. BenAdbelkader, R. Cutler, and L. Davis. Motion-based recognition of people using image self-similarity. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 254–259, 2002.
- [2] D. Cunado, M. S. Nixon, and J. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90:1–41, 2003.
- [3] A. Elgammal. Learning nonlinear manifolds for dynamic shape and dynamic appearance. Technical Report DCS TR 549, Department of Computer Science, Rutgers University, 2004.
- [4] A. Elgammal and C.-S. Lee. Separating style and content on a nonlinear manifold. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, June–July 2004.
- [5] J. B. Hayfron-Aquah, M. S. Nixon, and J. N. Carter. Automatic gait recognition by symmetry analysis. *Pattern Recognition Letters*, 24:2175–2183, 2003.
- [6] Q. He and C. Debrunner. Individual recognition from periodic activity using hidden markov models. In *In IEEE Workshop on Human Motion*, 2000.
- [7] P. Huang, C. Haris, and M. Nixon. Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13:359–366, 1999.
- [8] A. Y. Johnson and A. F. Bobick. A multi-view method for gait recognition using static body parameters. In *International Conference on Audio- and Video Based Biometric Person Authentication*, pages 301–311, June 2001.
- [9] A. Kale, A. Rajagopalan, N. Cuntoor, and V. Krueger. Human identification using gait. In *Proceedings FGR*, 2002.
- [10] L. Lee and W. Grimson. Gait analysis for recognition and classification. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 148–155, 2002.
- [11] J. J. Little and J. E. Boyd. Recognizing people by their gait: The shape of motion. *Videre: Journal of Computer Vision Research*, 1(2), 1998.
- [12] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 1988.
- [13] D. H. Marimont and B. A. Wandell. Linear models of surface and illuminant spectra. *Journal of Optical Society of America*, 9(11):1905–1913, 1992.
- [14] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.
- [15] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. Baseline results for the challenge problem of human id using gait analysis. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 137–142, 2002.
- [16] T. Poggio and F. Girosi. A theory of networks for approximation and learning. Technical Report A.I. Memo No.1140, MIT Artificial Intelligence Laboratory, 1989.
- [17] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [18] J. D. Shutler, M. S. Nixon, and C. J. Harris. Statistical gait description via temporal moments. In *IEE Colloquium-Visual Biometrics*, pages 11/1–11/5, 2000.
- [19] J. Tenenbaum. Mapping a manifold of perceptual observations. In *Advances in Neural Information Processing*, volume 10, pages 682–688, 1998.
- [20] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.