Automatic Gait Recognition by Multi-projection Analysis

Murat Ekinci

Computer Vision Lab.
Dept. of Computer Engineering,
Karadeniz Technical University, Trabzon, Turkey
`ekinci@ktu.edu.tr`

Abstract. Human identification at a distance by analysis of gait patterns extracted from video has recently become very popular research in biometrics. This paper presents multi-projections based approach to extract gait patterns for human recognition. Binarized silhouette of a motion object is represented by 1-D signals which are the basic image features called the distance vectors. The distance vectors are differences between the bounding box and silhouette, and extracted using four projections to silhouette. Based on normalized correlation on the distance vectors, gait cycle estimation is first performed to extract the gait cycle. Second, eigenspace transformation is applied to time-varying distance vectors and the statistical distance based supervised pattern classification is then performed in the lower-dimensional eigenspace for human identification. A fusion strategy developed is finally executed to produce final decision. Experimental results on four databases demonstrate that the right person in top two matches 100% of the times for the cases where training and testing sets corresponds to the same walking styles, and in top three-four matches 100% of the times for training and testing sets corresponds to the different walking styles.

1 Introduction

Gait recognition is the term typically used in the computer community to refer to the automatic extraction of visual cues that characterize the motion of a walking person in video and is used for identification purpose in surveillance [2][3][1][6]. Often in surveillance applications, it is difficult to get face information at the resolution required for recognition. As for gait is one of the few biometrics and a behavioral biometric source that can be measured at a distance.

Gait recognition methods can be broadly divided into two groups, model based and silhouette based approaches. Model based approaches [3][14] recover explicit features describing gait dynamics, such as stride dimensions and the kinematics of joint angles. The silhouette approach [6][7][8][2], characterizes body movement by the statistics of the patterns produced by walking. These patterns capture both the static and dynamic properties of body shape.

In this paper, it is attempted to develop a simple but effective representation of silhouette for gait-based human identification using silhouette analysis. Similar

observations have been made in [7][8][2], but the idea presented here implicitly captures both structural (appearances) and transitional (dynamics) characteristics of gait. A more robust approach for gait cycle estimation which is very important step in gait recognition is also presented. Instead of width/length time signal of bounding box of moving silhouette usually used in existing gait period analysis [3][9][10][2], here we analyze four projections extracted directly from differences between silhouette and the bounding box, and further convert them into associated four 1D signals. The novel approach presented is basically to produce the distance vectors, which are four 1-D signals are extracted for each projections, they are top-, bottom-, left-, and right-projections. Then normalized correlation-based a similarity function is executed to estimate gait cycle of moving silhouette. As following main purpose, depending on four distance vectors, PCA based gait recognition algorithm is first performed. A statistical distance based similarity is then achieved to obtain similarity measures on training and testing data. Next, fusion strategies on that similarities are calculated to produce final decision. Robust results for human identification have been obtained at the experiments on four different databases.

2   Gait Feature Extraction and Classification

Given a sequence of images obtained from a static camera, the moving person is first detected and tracked to compute the corresponding sequence of motion regions in each frame. Motion segmentation is achieved via background modeling/subtraction using a dynamic background frame estimated and updated in time, more details are given in [5]. Then a morphological filter operator is applied to the resulting images to suppress spurious pixel values. Once a silhouette generated, a bounding box is placed around silhouette. Silhouettes across a motion sequence are automatically aligned by scaling and cropping based on the bounding box.

2.1   Gait Signature Extraction

Silhouette representation is based on the projections to silhouette which is generated from a sequence of binary silhouette images $bs(t) = bs(x,y,t)$, indexed spatially by pixel location (x,y) and temporally by time t. There are four different image features called the distance vectors. They are top-, bottom-, left- and right-distance vectors. The distance vectors are the differences between the bounding box and the outer contour of silhouette. An example silhouette and the distance vectors corresponding to four projections are shown in the middle of figure 1. The distance vectors are separately represented by four 1D signals. The size of 1D signals is equal to the height or to the width of the bounding box for left- and right-distance vectors or for top- and bottom-distance vectors, respectively. The values in the signals for both left- and right-projections are computed as the difference in the locations of the bounding box and left-most and right-most boundary pixels, respectively, in a given row. The other projections along a given column are also computed as the differences from the top of
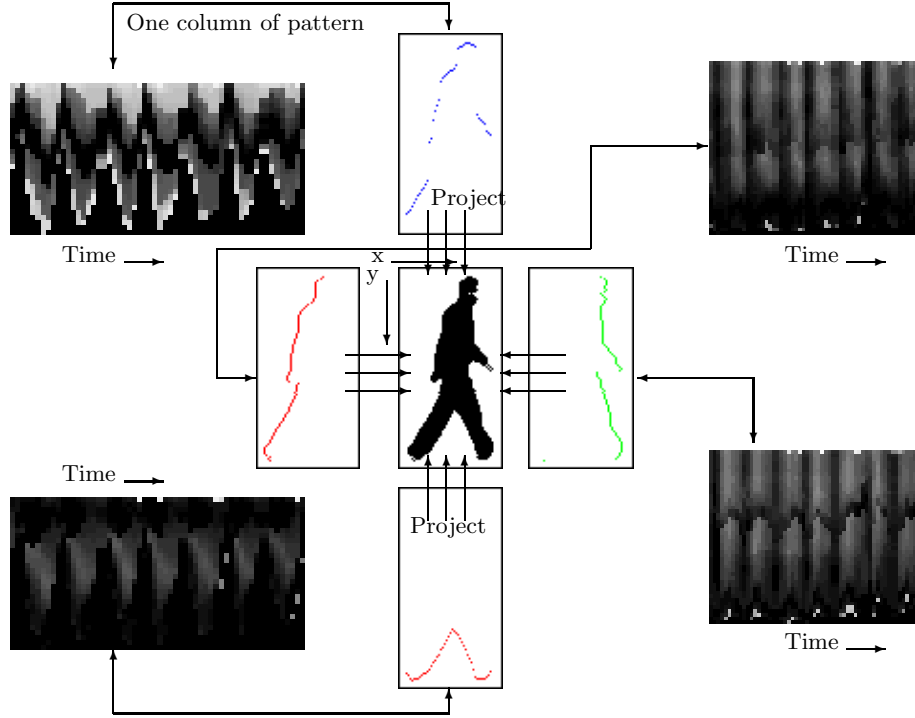
Fig. 1. Silhouette representation. **(Middle)** Silhouette and four projections, **(Left)** temporal plot of the distance vectors for top and bottom projections, **(Right)** temporal plot of the distance vectors for left and right projections.

the bounding box to the top-most of silhouette for top-projection, from the bottom of the box to the bottom-most of silhouette pixels for bottom-projections, respectively.

From a new 2D image $F_T(x,t) = \sum_y bs(x,y,t)$, where each column (indexed by time t) is the top-projections (row sum) of silhouette image $bs(t)$, as shown in figure 1 top-left. Each value $F_T(x,t)$ is then a count of the number of the row pixels between the top side of the bounding box and the outer contours in that columns x of silhouette image $bs(t)$. The result is a 2D pattern, formed by stacking top projections together to form a spatio-temporal pattern. A second pattern $F_B(x,t) = \sum_{-y} bs(x,y,t)$ can be constructed by stacking bottom projections, as shown in figure 1 bottom-left. The third pattern $F_L(y,t) = \sum_x bs(x,y,t)$ is also constructed by stacking with using the differences produced from the left projections and the last pattern $F_R(y,t) = \sum_{-x} bs(x,y,t)$ is finally constructed by stacking the right projections, as shown in figure 1 top-right and bottom-right 2D patterns, respectively. The variation of each component of the distance vectors can be regarded as gait signature of that object. From the temporal distance vector plots, it is clear that the distance vector is roughly periodic and gives the

extent of movement of different part of the subject. The brighter a pixel in 2D patterns in figure 1, the larger value is the value of the distance vector in that position.

2.2   Training

The representation of 2D silhouette shape by four 1D signals, called distance vectors, significantly reduces the subsequent computational cost. To eliminate the influence of spatial scale and signal length of the distance vectors, the algorithm scales these distance vector signals with respect to magnitude and size through the sizes of the bounding boxes. Next, eigenspace transformation based on Principal Component Analysis (PCA) is applied to time varying distance vectors derived from a sequence of silhouette images to reduce the dimensionality of the input feature space. The training process similar to [2][4] is illustrated as follows:

Given k class for training, and each class represents a sequence of the distance vector signals of a person in one gait cycle. Multiple sequences of each person can be separately added for training. Let $V_{i,j}^w$ be the jth distance vector signal in the ith class for w projection to silhouette and N_i the number of such distance vector signals in the ith class. The total number of training samples is $N_t^w = N_1^w + N_2^w + ... + N_k^w$, as the whole training set can be represented by $[V_{1,1}^w, V_{1,2}^w, .., V_{1,N_1}^w, V_{2,1}^w, ..., V_{k,N_k}^w]$. The mean m_v^w and the global covariance matrix $\sum^w$ of w projection training set can easily be obtained by

$$m_v^w = \frac{1}{N_t^w} \sum_{i=1}^{k} \sum_{j=1}^{N_i^w} V_{i,j}^w \tag{1}$$

$$\sum{}^w = \frac{1}{N_t^w} \sum_{i=1}^{k} \sum_{j=1}^{N_i^w} (V_{i,j}^w - m_v^w)(V_{i,j}^w - m_v^w)^T \tag{2}$$

Here each V^w represents the distance vectors, F_w, for w projection (Top-Bottom-Left-Right) as explained in section 2.1. If the rank of matrix $\sum$ is N, then the N nonzero eigenvalues of $\sum$, $\lambda_1, \lambda_2, .., \lambda_N$, and associated eigenvectors $e_1, e_2, .., e_N$ can be computed based on theory of *singular value decomposition* [4]. The first few eigenvectors correspond to large changes in training patterns, and higher-order eigenvectors represent smaller changes [2]. As a result, for computing efficiency in practical applications, those small eigenvalues and their corresponding eigenvectors are ignored. Then a transform matrix $T^w = [e_1^w, e_2^w, .., e_s^w]$ to project an original distance vector signal $V_{i,j}^w$ into a point $P_{i,j}^w$ in the eigenspace is constructed by taking only $s < N$ largest eigenvalues and their associated eigenvectors for each projections to silhouette. Therefore, s values are usually much smaller than the original data dimension N. Then the projection average $A_i{}^w$ of each training sequence in the eigenspace is calculated by averaging of $P_{i,j}^w$.

2.3   Pattern Classification

Statistical distance. measuring has initially been selected for classification. The accumulated distance between the associated centroids A^w (obtained in the process of training) and B^w (obtained in the process of testing) can be easily computed by

$$d_S(A, B) = \sqrt{(\frac{A_1 - B_1}{s_1})^2 + ... + (\frac{A_p - B_p}{s_p})^2} = \sqrt{(A - B)^t S^{-1} (A - B)} \quad (3)$$

Where $S = diag(s_1^2, ..., s_p^2)$. In the distance measure, the classification result for each projection is then accomplished by choosing the minimum of d. The classification process is carried out via the nearest neighbor classifier. The classification is performed by classifying the test sequence into class c that can minimize the similarity distance between the test sequence and all training sequences by

$$c = arg_i \ min \ d_i(T, R_i) \quad (4)$$

where T represents a test sequence, R_i represents the ith training sequence, d is the similarity measures described in above.

A fusion task includes two different strategies is finally developed. In **the strategy 1**, each projection is separately treated. Then the strategy 1 is combining the distances of each projections at the end by assigning equal weight. The final similarity using strategy 1 is calculated as follows:

$$D_i = \sum_{j=1}^{4} \alpha_j * d_{ji} \quad (5)$$

where D_i is the fused distance similarity value, j is the algorithm's index for projection, α its normalized weight, d_i its single projection distance similarity value and 4 is the number of projections (left, right, top, bottom). As conclusion, if any two distance similarity vectors in four projections give maximum similarities for same person, then the identification is determined as positive. This fusion strategy has rapidly increased the recognition performance in the experiments.

At the experiments, it has been seen that some projection has given more robust results than others. For example, while human moves in lateral view with respect to image plane, the back side of human can give more individual characteristics in gait. So, the projection corresponding to that side can give more reliable results. We called dominant feature to this case. As a result, **the strategy 2** has also been developed to further increase the recognition performance. In the strategy 2, if the projection selected as dominant feature gives positive or at least two projections of others give positive for an individual, then identification by fusion process is to be positive. The dominant feature in this work is automatically assigned by estimating the direction of motion objects under tracking. At the next section, the dominant features determined by experimentally for different view points with respect to image plane are also given.

View 1   View 2   View 3   View 4   View 5   View 6

Fig. 2. The six CMU database viewpoints

3   Experiments and Results

We evaluate the performance of the method on CMU's MoBo database[12], NLPR gait database [2], our database, and USF database [6]. The Viterbi algorithm was used to identify the test sequence, since it is efficient and can operate in the logarithmic domain using only additions [11]. The performance of the algorithm presented is evaluated on four different databases of varying of difficulty. The gait cycle is automatically estimated by using the algorithm presented in the previous work in [13].

CMU Database. This database has 25 subjects (23 males, 2 females) walking on a treadmill. Each subject is recorded performing four different types of walking: slow walk, fast walk, inclined walk, and slow walk holding ball. There are about 8 cycles in each sequence, and each sequences is recorded at 30 frames per second. It also contains six simultaneous motion sequences of 25 subjects, as shown in figure 2.

One of the cycle in each sequence was used for testing, others for training. First, we did the following experiments on this database: **1)** train on slow walk and test on slow walk, **2)** train on fast walk and test on fast walk, **3)** train on walk carrying a ball and test on walk carrying a ball, **4)** train on slow walk and test on fast walk, **5)** train on slow walk and test on walk carrying a ball, **6)**

Table 1. Classification performance on the CMU data set for viewpoint 1

Test/Train	All projections: equal				Dominant: Right projection		
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 1	Rank 2	Rank 3
Slow/Slow	72	100	100	100	84	100	100
Fast/Fast	76	100	100	100	92	100	100
Ball/Ball	84	100	100	100	84	100	100
Slow/Fast	36	92	100	100	52	100	100
Fast/Slow	20	60	100	100	32	88	100
Slow/Ball	8	17	33	58	42	96	100
Fast/Ball	4	13	33	67	17	50	88
Ball/Slow	8	17	38	67	33	88	100
Ball/Fast	13	29	58	92	29	63	100

Table 2. Classification performance on the CMU data set for all views. Eight gait cycles were used, seven cycles for training, one cycle for testing.

View	Test/Train	All projections equal			Dominant: Right projection		
		Rank 1	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
4	Slow/Slow	76	100	100	84	100	100
	Fast/Fast	84	100	100	96	100	100
	Slow/Fast	12	44	80	24	64	100
	Fast/Slow	20	64	100	32	76	100
					Dominant: Left projection		
5	Slow/Slow	80	100	100	80	100	100
	Fast/Fast	88	100	100	88	100	100
	Slow/Fast	16	44	80	24	64	100
	Fast/Slow	24	56	96	32	68	100
					Dominant: Right projection		
3	Slow/Slow	80	100	100	88	100	100
	Fast/Fast	72	100	100	76	100	100
	Slow/Fast	20	64	100	28	76	100
	Fast/Slow	24	56	92	28	68	100
					Dominant: Right projection		
6	Slow/Slow	72	100	100	84	100	100
	Fast/Fast	76	100	100	80	100	100
	Slow/Fast	16	44	88	36	76	100
	Fast/Slow	16	40	72	24	56	100

train on fast walk and test on slow walk, **7)** train on fast walk and test on walk carrying a ball, **8)** train on walk carrying a ball and test on slow walk, **9)** train on walk carrying a ball and test on fast walk.

The results obtained using the proposed method are summarized on the all cases **1)-9)** in Table 1. It can be seen that the right person in the top two matches 100% of times for the cases where testing and training sets correspond to the same walk styles. When the strategy developed in the fusion as dominant feature (projections) is used, the recognition performance is increased, as seen in Table 1. For the case of training with fast walk and testing on slow walk, and vice versa, the dip in performance is caused due to the fact that for some individual as biometrics suggests, there is a considerable change in body dynamics and stride length as a person changes his speed. Nevertheless, the right person in the top three matches 100% of the times for that cases, and dominant projection strategy has also increased the recognition performance for Ranks 1 and 2. For the case of training with walk carrying ball and testing on slow and fast walks, and vice versa, encouraging results have also been produced by using the proposed method, and the dominant feature property has still increased the recognition performance, as given in Table 1.

For the other view points, the experimental results are also summarized on the cases **1)-4)** in Table 2. When the all experimental results for the different view points are considered, it can be seen that, the right person in the top two matches 100% and in the top four matches 100% of the times for the cases **1)-2)**
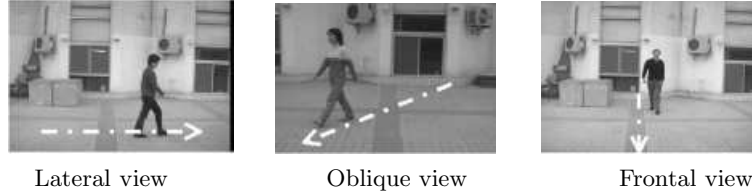
Lateral view   Oblique view   Frontal view

Fig. 3. Some images in the NLPR database

and for the cases **3)-4)**, respectively. It is also seen that, when the dominant feature is used, gait recognition performance is also increased. Consequently, the proposed method for gait recognition can easily be seen that, it is view-invariant.

NLPR Database. The *NLPR* database [2] includes 20 subjects and four sequences for each viewing angle per subject, two sequences for one direction of walking, the other two sequences for reverse direction of walking. For instance, when the subject is walking laterally to the camera, the direction of walking is from right to left for two of four sequences, and from right to left for the remaining. Those all gait sequences were captured as twice (we called two experiments) on two different days in an outdoor environment. All subjects walk along a straight-line path at free cadences in three different views with respect to the image plane, as shown in figure 3, where the white line with arrow represents one direction path, the other walking path is reverse direction.

We did the following experiments on this database: **1)** train on one image sequence and test on the remainder, all sequences were produced from first experiment, **2)** train on two sequences obtained from first experiment and test on two sequences obtained from second experiment. This is repeated for each viewing angle, and for each direction of walking. The results for the experiments along with cumulative match scores in three viewing angle are summarized in Table 3. When the experimental results are considered, the right person in the

Table 3. Performance on the NLPR data set for three views

Walking Direction	View	Training	Test	Rank1	Rank2	Rank3
One Way Walking	Lateral	Exp. 1	Exp. 1	65	100	100
		Exp. 1	Exp. 2	55	100	100
	Frontal	Exp. 1	Exp. 1	60	100	100
		Exp. 1	Exp. 2	35	100	100
	Oblique	Exp. 1	Exp. 1	40	90	100
		Exp. 1	Exp. 2	30	60	100
Reverse Way Walking	Lateral	Exp. 1	Exp. 1	60	100	100
		Exp. 1	Exp. 2	50	100	100
	Frontal	Exp. 1	Exp. 1	60	100	100
		Exp. 1	Exp. 2	40	100	100
	Oblique	Exp. 1	Exp. 1	45	100	100
		Exp. 1	Exp. 2	35	75	100

Table 4. Performance on our data set: (The abbreviation: L⇒ R : From Left to Right)

Direction	Outdoor(15 person)		Indoor(7 person)		All(22 person)		
	Rank 1	Rank 2	Rank 1	Rank 2	Rank 1	Rank 2	Rank 3
L⇒ R	67	100	86	100	68	95	100
R⇒ L	67	100	71	100	68	100	100

top two matches 100% of times for lateral and frontal viewing angles, and in the top three matches 100% of times for oblique view.

Our Database. The database established for gait recognition has 22 people (2 females, 20 males), and subjects are walking laterally to the camera, the directions of walking is from left to right, and from right to left. The database includes two sequences for each subject. One sequence includes 3 gait cycle for each direction, and the length of each gait cycle varies with the pace of the walker, but the average is about 26 frames. The subjects walk along a straight-line path at free cadences, and 15 subjects were walking outside, seven subjects were walking inside. The results for the experiments along with cumulative match scores in lateral view are also summarized in Table 4. Three gait cycles were used, two cycles for training, one cycle for testing. Walking from left to right and the other direction are separately tested to achieve initial experimental results. When the results of each projection based distance vectors are re-implemented by using dominant feature strategy as explained in section 2.3, significantly improvements on the gait recognition has also been achieved. This is the robust implementation and is one of the novelty presented by this paper.

USF Database. Finally, the USF database [6] is considered. The database has variations as regards viewing direction, shoe type, and surface type. At the experiments, one of the cycle in each sequence was used for testing, others (3-4 cycles) for training. Different probe sequences for the experiments along with the cumulative match scores are given in Table 5 for the algorithm presented in this paper and three different algorithms [15][2][7]. The same silhouette data from USF were directly used. These data are noisy, e.g., missing of body parts, small holes inside the objects, severe shadow around feet, and missing and adding

Table 5. Performance on the USF database for four algorithm.

Exp.	The method			Baseline[15]		NLPR[2]		UMD[7]	
	Rnk 1	Rnk 2	Rnk 3	Rnk 1	Rnk 5	Rnk 1	Rnk 5	Rnk 1	Rnk 5
GAL[68]	35	80	100	79	96	70	92	91	100
GBR[44]	34	82	100	66	81	58	82	76	81
GBL[44]	25	55	91	56	76	51	70	65	76
CAL[68]	39	90	100	30	46	27	38	24	46
CAR[68]	30	66	100	29	61	34	64	25	61
CBL[41]	30	78	100	10	33	14	26	15	33
CBR[41]	29	66	100	24	55	21	45	29	39
GAR[68]	34	60	90	−	-	-	-	-	-

some parts around the border of silhouettes due to background characteristics. In Table 5, G and C indicate grass and concrete surfaces, A and B indicate shoe types, and L and R indicate left and right cameras, respectively. The number of subjects in each subset is also given in square bracket. It is observed that, the proposed method has given the right person in top three matches 100% of the times for training and testing sets corresponding to the same camera.

4   Conclusion

This paper has described a novel gait recognition approach that uses multi projections of silhouettes as the basic feature for classification. The approach on gait cycle estimation and on gait recognition is view-invariant and is fully automatic for real time applications. The performance of the gait recognition method proposed was also illustrated using different gait databases.

References

1. G. V. Veres,*et.al*, What image information is important in silhouette-based gait recognition? Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2004.
2. L. Wang, T. Tan, H. Ning, W. Hu, Silhouette Analysis-Based Gait Recognition for Human Identification. IEEE Trans. on PAMI Vol.25, No. 12, Dec.,2003.
3. C. BenAbdelkader, R. G. Cutler, L. S. Davis, Gait Recognition Using Image Self-Similarity. EURASIP Journal of Applied Signal Processing, April, 2004.
4. P. Huang, C. Harris, M. Nixon, Human Gait Recognition in Canonical Space Using Temporal Templates. IEE Proc. Vision Image and Signal Proc. Conf., 1999.
5. M. Ekinci, E. Gedikli, Background Estimation Based People Detection and Tracking for Video Surveillance. Springer LNCS Vol. 2869, pp:421-429, November, 2003.
6. S. Sarkar, *et al* The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. IEEE Trans. on Pat. Anal. and Mach. Intell., Vol.27, No. 2, 2005.
7. A. Kale, *et. al.*, Identification of Humans Using Gait. IEEE Trans. on Image Processing, Vol.13, No.9, September 2004.
8. Yanxi Liu, R. T. Collins, T. Tsin, Gait Sequence Analysis using Frieze Patterns. Proc. of European Conf. on Computer Vision, 2002.
9. C. BenAbdelkader, *et.al*, Stride and Cadence as a Biometric in Automatic Person Identification and Verification. Proc. Int. Conf. Aut. Face and Gesture Recog.,2002.
10. R. Collins, R. Gross, and J. Shi, Silhouette-Based Human Identification from Body Shape and Gait. Proc. Int. Conf. Automatic Face and Gesture Recognition, 2002.
11. J. Phillips *et.al*, The FERET Evaluation Methodology for Face recognition Algorithm. IEEE Trans. Pattern Analysis and Machine Intell. Vol.22, no.10, Oct.2000.
12. R. Gross, J. Shi, The CMU motion of body (MOBO) database. Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.
13. M. Ekinci, E. Gedikli A Novel Approach on Silhouette Based Human Motion Analysis for Gait Recognition. Springer Verlag LNCS Vol.3804, pp.219-226, Dec.2005.
14. A. I. Bazin, M. S. Nixon, Gait Verification Using Probabilistic Methods. IEEE Workshop on Applications of Computer Vision, 2005.
15. P. Phillips, et.al., Baseline Results for Challenge Problem of Human ID using Gait Analysis. Pro. Int. Conf. Automatic Face and Gesture Recognition, 2002.