A New Approach for Human Identification Using Gait Recognition

Murat Ekinci

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

Abstract. Recognition of a person from gait is a biometric of increasing interest. This paper presents a new approach on silhouette representation to extract gait patterns for human recognition. Silhouette shape of a motion object is first represented by four 1-D signals which are the basic image features called the distance vectors. The distance vectors are differences between the bounding box and silhouette. Second, eigenspace transformation based on Principal Component Analysis is applied to time-varying distance vectors and the statistical distance based supervised pattern classification is then performed in the lower-dimensional eigenspace for recognition. A fusion task is finally executed to produce final decision. Experimental results on three databases show that the proposed method is an effective and efficient gait representation for human identification, and the proposed approach achieves highly competitive performance with respect to the published gait recognition approaches.

1 Introduction

Human identification from gait has been a recent focus in computer vision. It is a behavioral biometric source that can be acquired at a distance. Gait recognition aims to discriminate individuals by the way they walk and has the advantage of being non-invasive, hard to conceal, being readily captured without a walker's attention, and is less likely to be obscured than other biometric features [1][2][3][6].

Gait recognition can be broadly divided into two groups, model-based and silhouette-based methods. Model-based methods [2][15] model the human body structure and extract image features to map them into structural components of models or to derive motion trajectories of body parts. The silhouette-based methods [6][7][9][1], 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 an effective representation of silhouette for gait recognition is developed and statistical analysis is performed. Similar observations have been made in [7][9][1], but the idea presented here implicitly captures both structural (appearances) and transitional (dynamics) characteristics of gait. The silhouette-based method presented is basically to produce the distance vectors, which are four 1D signals extracted from projections to silhouette, they are top-, bottom-,

M. Gavrilova et al. (Eds.): ICCSA 2006, LNCS 3982, pp. 1216–1225, 2006.

left-, and right-projections. 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. A fusion task includes two strategies is executed to produce consolidation decision. Experimental results on three different databases demonstrate that the proposed algorithm has an encouraging recognition performance.

2 Silhouette Representation

To extract spatial silhouettes of walking figures, a background modeling algorithm and a simple correspondence procedure are first used to segment and track the moving silhouettes of a walking figure, more details are given in [5]. Once a silhouette generated, a bounding box is placed around silhouette. Silhouette across a motion sequence (a gait cycle) are automatically aligned by scaling and cropping based on the bounding box. The details on the gait cycle estimation used are given in reference [14].

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 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 which represents the bottom-projection $F^B(x,t) = \sum_{-y} bs(x,y,t)$ 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 then constructed by stacking with using the differences as column pixels from left side of the box to left-most boundary pixels of silhouette which are produced by the left projections, and the last pattern $F^R(y,t) = \sum_{-x} bs(x,y,t)$ is also finally constructed by stacking the right projections, as shown in figure 1 top-right and bottom-right 2D patterns,

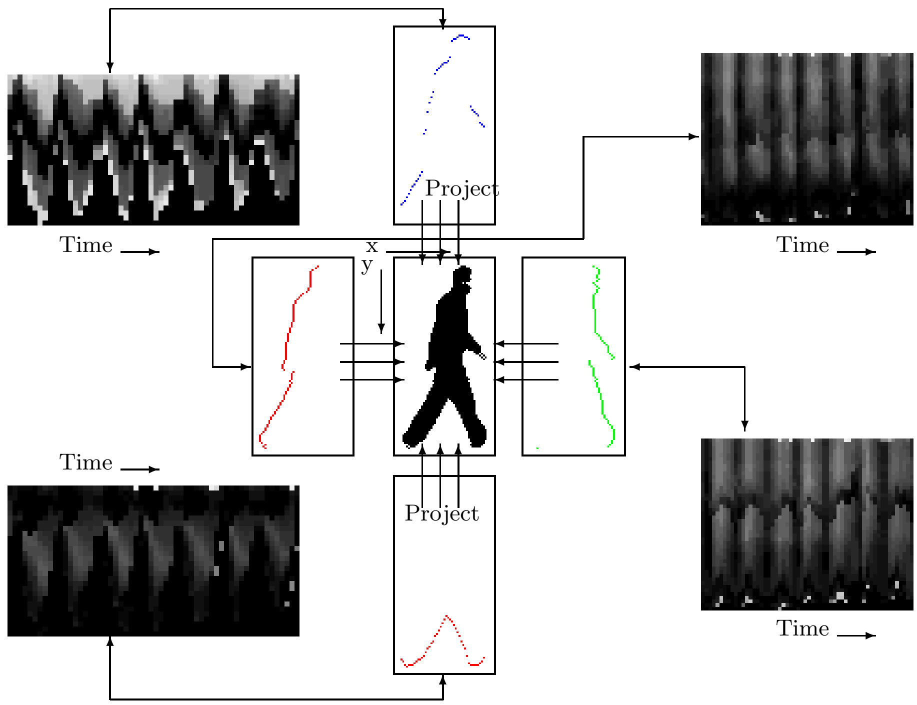


Fig. 1. Silhouette representation. **(Middle)** Silhouette and four projections, **(Left)** temporal plot of the distance vectors for top and bottom projections, **(Right)** temporal plot of the distance vectors for left and right projections.

respectively. The variation of each component of the each 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.

3 Training

The following processes on the four 1D signals produced from the distance vectors are to eliminate the influence of spatial scale and signal length of the distance vectors by scaling of these distance vector signals with respect to magnitude and size through the sizes of the bounding boxes. Eigenspace transformation based on Principal Component Analysis (PCA) is then 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 [1][4] is illustrated as follows:

Given k class for training, and each class represents a sequence of the distance vector signals of a person. Multiple sequences of each subject can be added for training, but a sequence includes one gait cycle was considered in the experiments. Let $V_{i,j}^w$ be the *j*th distance vector signal in the *i*th class for w projection to silhouette and N_i the number of such distance vector signals in the *i*th 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 value is 1D signal and equal to, $F^w(.)$, the distance vectors for w projection (top-bottom-left-right) as explained in section 2. 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 [1]. 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 $P_{i,j}^w$ as follows:

$$P_{i,j}^w = [e_1^w \;\; e_2^w \;\; ... \;\; e_s^w]^T V_{i,j}^w, \qquad A_i^w = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{i,j}^w \tag{3}$$

4 Pattern Classification

Gait pattern recognition (classification) can be solved through measuring similarities between reference gait pattern and test samples in the parametric eigenspace. A simple statistical distance has been chosen to measure similarity, because the main interest here is to evaluate the genuine discriminatory ability of the extracted features in the proposed method. 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

$$d_S(A, B) = \sqrt{(\frac{A_1 - B_1}{s_1})^2 + ... + (\frac{A_p - B_p}{s_p})^2} \tag{4}$$

Where $(s_1, ..., s_p)$ are equal to corresponding the sizes of A_i and B_i. 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 (NN) classifier. The classification is performed by classifying in the test sequence and all training sequences by

$$c = \arg_i \ min \ d_i(B, A_i) \tag{5}$$

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

The similarity results produced from each distance vectors are fused to increase the recognition performance. In this fusion task, two different strategies were developed. In **the strategy 1**, each projection is separately treated. Then the strategy is combining the distances of each projections at the end by assigning equal weight. As implementation, if any two of the similarities achieved based on four projections give maximum similarities for same individual, then the identification is appointed 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 instance, 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. It is called dominant feature to this case. As second strategy, **the strategy 2** has also been developed to further increase the recognition performance. In the strategy 2, if the projection selected as dominant feature or at least two projections of others give positive for an individual, then identification result given by the strategy 2 is appointed as positive. The dominant feature in this work is automatically assigned by estimating the direction of motion objects in tracking. At the next section, the dominant features determined by experimentally for different view points with respect to image plane are given.

5 Experimental Results

The performance of the proposed methods was evaluated on CMU's MoBo database[13], NLPR gait database [1], 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 [12]. The performance of the algorithm is evaluated on three different databases of varying of difficulty.

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 sequence 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

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

	All projections: equal				Dominant: Right projection		
Test/Train	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

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)** 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.

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

Fig. 2. The six CMU database viewpoints

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.

		All projections equal			Dominant: Right projection		
View	Test/Train	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

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)** and for the cases **3)-4)**, respectively. It is also seen that, when the dominant feature is used, gait recognition performance is also increased. Some comparison results are also given in Table 3. The reason to show the cases **1)-6)** only is become the page limitation in this paper, and that points are our lowest results on MoBo dataset for the comparisons to the other works in literature.

NLPR Database. The *NLPR* database [1] includes 20 subjects and four sequences for each viewing angle per subject, two sequences for one direction of

Table 3. Comparison of several algorithm on MoBo dataset

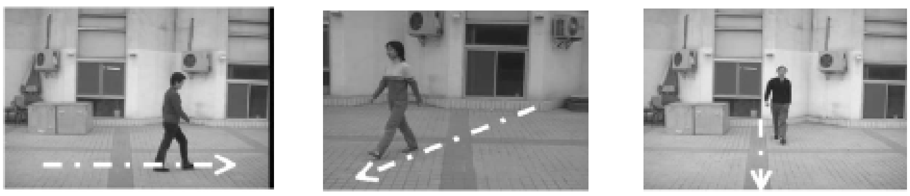
Algorithms	train/test	Rank 1(%)	Rank 2(%)	Rank 3 (%)	Rank 5 (%)
The method presented	slow/slow	84	100	100	100
Kale *et.al.* [7]	slow/slow	72	80	85	97
Collins *et.al* [11]	slow/slow	86	100	100	100
The method presented	fast/slow	52	100	100	100
Kale *et.al.* [7]	fast/slow	56	62	75	82
Collins *et.al.*[11]	fast/slow	76	Not	given	92

Table 4. 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

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 4. When the experimental results are considered, the right person in the top two matches 100% of times for lateral and frontal viewing angles, and in the top three matches 100% of times for oblique view.



Lateral view Oblique view Frontal view

Fig. 3. Some images in the NLPR database

In the experiments on the NLPR database, the performance of the proposed algorithm was also compared with those of a few recent silhouette-based methods described in [11],[8],[16], and [1], respectively. To some extent, they reflect the latest and best work of these research groups in gait recognition. In [11], a method based on template matching of body silhouettes in key frames for human identification was established. The study in [8] described a moment-based representation of gait appearance for the purpose of person identification. A baseline algorithm was also proposed for human identification using spatio temporal correlation of silhouette images in [16]. The work in [1] computes the centroid of silhouette's shape, and unwraps the outer counter to obtain a 1D distance signal, then applies principal component analysis for person identification. These methods were implemented using the same silhouette data from the NLPR database with lateral view by the study in [1], and the results given are as taken from tables in [1]. Table 5 lists the identification rates that have been reported by other algorithms and our algorithm. The proposed algorithm has successfully given the right person in top two matches 100% the times for the NLPR database.

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 6 for the algorithm presented in this paper and three different algorithms [16][1][7]. The same silhouette data from USF were directly used. These data are noisy, e.g., missing of body parts, small

Table 5. Comparison of Several algorithm on the NLPR Database (Lateral View)

Methods	Rank 1(%)	Rank 2(%)	Rank 3 (%)	Rank 5 (%)	Rank 10 (%)
BenAbdelkader [10]	73	Not	given	89	96
Collins [11]	71	Not	given	79	88
Lee [8]	88	Not	given	99	100
Phillips [16]	79	Not	given	91	99
Wang [1]	75	Not	given	98	100
The methods presented	65	100	100	100	100

Table 6. Performance on the USF database for four algorithm

	The method			Baseline[16]		NLPR[1]		UMD[7]	
Exp.	Rank 1	Rank 2	Rank 3	Rank 1	Rank 5	Rank 1	Rank 5	Rank 1	Rank 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	–	-	-	-	-	-

holes inside the objects, severe shadow around feet, and missing and adding some parts around the border of silhouettes due to background characteristics. In Table 6, 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.

6 Conclusion

The method presented has given very close results to the existing works for Rank 1 on the databases tested, nevertheless it has almost given 100% accuracy for Ranks 2 and 3 on the all databases used. Nonlinear discriminant analysis will be developed as next study to achieve higher accuracy than the current for Rank 1.

References

1. 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.
2. C. BenAbdelkader, R. G. Cutler, L. S. Davis, Gait Recognition Using Image Self-Similarity. EURASIP Journal of Applied Signal Processing, April, 2004.
3. 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.
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 2869, ISCIS 2003, 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. L. Lee, W. Grimson, Gait Analysis for Recognition and Classification. Proc. IEEE, Int. Conference on Automatic Face and Gesture Recognition, pp. 155-162, 2002.
9. Yanxi Liu, R. T. Collins, T. Tsin, Gait Sequence Analysis using Frieze Patterns. Proc. of European Conf. on Computer Vision, 2002.
10. 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.
11. 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.
12. 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.
13. 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.
14. M. Ekinci, E. Gedikli A Novel Approach on Silhouette Based Human Motion Analysis for Gait Recognition. ISVC 2005, LNCS 3804, pp.219-226, December 2005.
15. A. I. Bazin, M. S. Nixon, Gait Verification Using Probabilistic Methods. IEEE Workshop on Applications of Computer Vision, 2005.
16. P. Phillips, et.al., Baseline Results for Challenge Problem of Human ID using Gait Analysis. Proc. Int. Conf. Automatic Face and Gesture Recognition, 2002.