

Human Identification Using Gait

Murat EKİNCİ

Computer Vision Lab.

Dept. of Computer Engineering

Karadeniz Technical University, Trabzon, TURKEY

e-mail: ekinci@ktu.edu.tr

Abstract

Gait refers to the style of walking of an individual. This paper presents a view-invariant approach for human identification at a distance, using gait recognition. Recognition of a person from their gait is a biometric of increasing interest. Based on principal component analysis (PCA), this paper describes a simple, but efficient approach to gait recognition. Binarized silhouettes of a motion object are represented by 1-D signals, which are the basic image features called distance vectors. The distance vectors are differences between the bounding box and silhouette, and are extracted using 4 projections of the silhouette. Based on normalized correlation of the distance vectors, gait cycle estimation is first performed to extract the gait cycle. Second, eigenspace transformation, based on PCA, is applied to time-varying distance vectors and Mahalanobis distances-based supervised pattern classification are then performed in the lower-dimensional eigenspace for human identification. A fusion strategy is finally executed to produce a final decision. Experimental results on 3 main databases demonstrate that the right person in the top 2 matches 100% of the time for the cases where training and testing sets corresponds to the same walking styles, and in the top 3-4 matches 100% of the time when training and testing sets do not correspond to the same walking styles.

Key Words: *Gait recognition, biometric identification, computer vision, pattern recognition.*

1. Introduction

Gait is a behavioral biometric source that can be acquired at a distance. 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 purposes in surveillance systems [1-5, 8, 13]. Often in surveillance applications, it is difficult to get face or iris information at the resolution required for recognition. Studies in psychophysics [10] indicate that humans have the capability of recognizing people from even impoverished displays of gait, indicating the presence of identity information in gait.

1.1. Overview and contribution of the proposed method

The aim of this work is to establish an automatic gait recognition method based upon spatio-temporal silhouette analysis measured during walking. Gait includes both the body appearance and dynamics of

human walking motion [11]. Recognizing people by gait intuitively depends on how the silhouette shape of an individual changes over time in an image sequence.

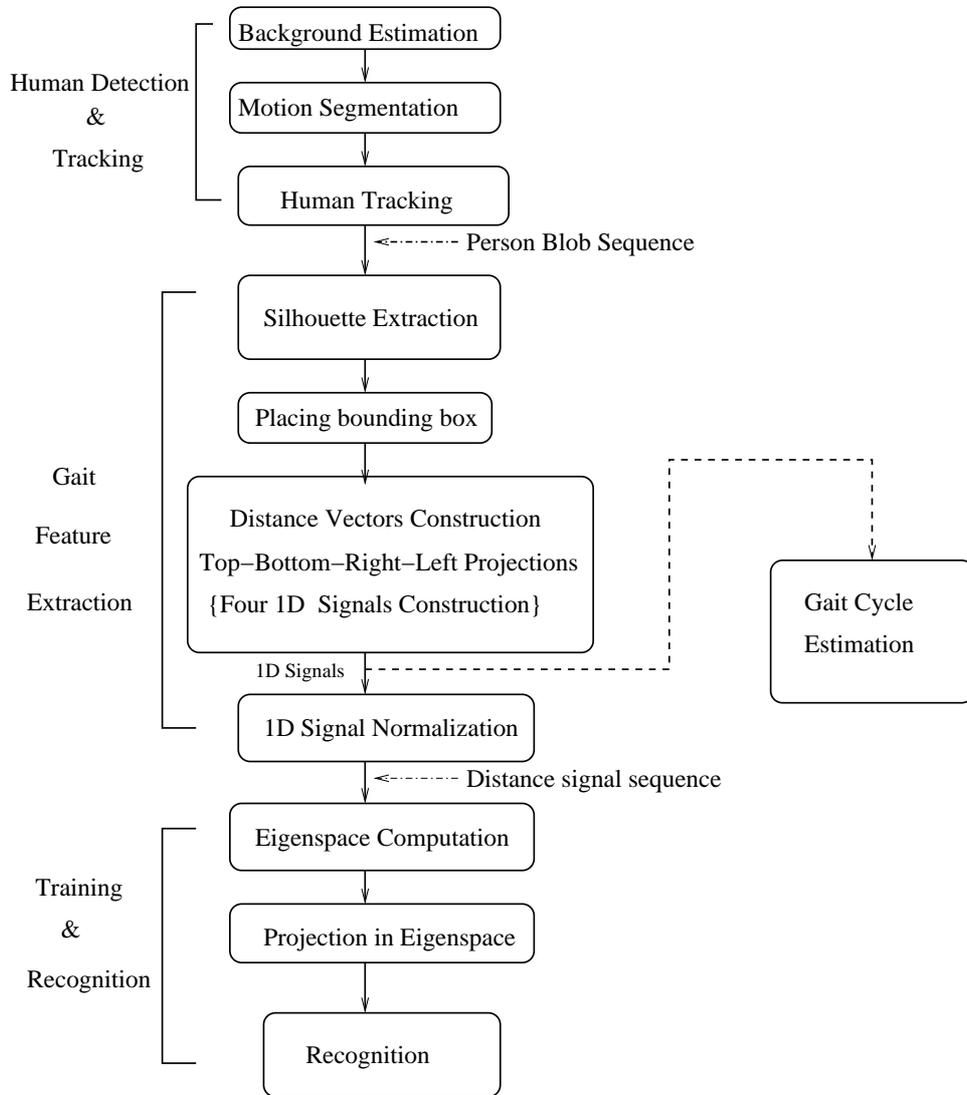


Figure 1. Overview of the proposed algorithm.

The overview of the proposed algorithm is shown in Figure 1. It consists of 3 major parts, namely, human detection and tracking, feature extraction, and training and classification. The first part serves to detect and track the walking figure in an image sequence. A background estimation procedure is performed to separate motion from the background, and the moving region corresponding to the spatial silhouette of the walking figure is successively tracked through a simple correspondence method [17]. The second part is used to extract the binary silhouette from each frame and map the 2D silhouette image into four 1D distance signals by producing 4 projections of the silhouette. Accordingly, the shape changes of these silhouettes over time are transformed into 4 sequences of four 1D distance signals to approximate temporal changes of gait pattern. These 1D signals are also used to estimate gait cycle, which is very important for

gait identification. The third part either applies principal component analysis (PCA) on those time-varying distance signals to compute the predominant components of gait signatures (training), or determines the person's identity using standard non-parametric pattern classification techniques in the lower-dimensional eigenspace (classification).

The main purpose and contributions of this paper are summarized as follows:

- We attempted to develop a simple, but effective representation of silhouette for gait-based human identification using silhouette analysis. Similar observations have been made in [3, 9, 14] but the idea presented here implicitly captures more structural (appearances) and transitional (dynamics) characteristics of gait.
- Instead of the width/length time signal of the bounding box of a moving silhouette usually used in existing gait period analysis [3, 4, 15, 16] here, we analyze 4 projections extracted directly from differences between the silhouette and the bounding box, and further convert them into associated four 1D signals.
- Instead of one silhouette image, which is usually used in existing silhouette-based work, here, 4 projections of a silhouette are analyzed and further are converted into associated four 1D signals. This greatly reduces the computational cost of the subsequent processes.

The presented method is integrated into a low-cost PC-based real-time view invariant visual surveillance system for gait cycle estimation and human motion analysis; these activities are in monochromatic video and then adapted into initial studies on gait recognition for human identification. This novel approach is basically to produce distance vectors, which are four 1-D signals extracted for each projection; top-, bottom-, left-, and right-projections. After they are normalized and correlation-based, a similarity function is executed to estimate the gait cycle of the moving silhouette. Following the main purpose, depending on the 4 distance vectors, a PCA-based gait recognition algorithm is first performed. Mahalanobis distances-based similarity is then achieved to obtain similarity measures on training and testing data. Next, fusion strategies on the similarities are calculated to produce a final identification decision. More robust results for human identification have been obtained with the experiments on 3 different databases. The proposed algorithm presents view-invariant opportunities in order to obtain the motion data and individual characteristics of silhouettes for gait recognition. The view-invariant characteristics of the algorithm have been exhibited by testing different video sequences produced from 3 different views, with respect to the image plane.

The remainder of this paper is organized as follows: Section 2 describes related work in the literature so as to put ours in context; Section 3 introduces feature extraction, which involves silhouette shape representation, gait cycle estimation, and the training process based on PCA, etc; Gait recognition based on the standard pattern classification technique is discussed in Section 4; Experimental results are also presented in Section 5; Finally, conclusions are provided in Section 6.

2. Related Works

Interest in gait recognition is best evidenced by the near-exponential growth of the related literature during the past few years [1-4, 8, 9, 12, 14, 24] Gait recognition includes human motion analysis methods that automatically detect and track human motions, then gait cycle estimations, next, the extraction of the characteristics of an individual, and finally, recognition methods for human identification.

The study in [24] extracts 2D XT sheets that encode the person's inner and outer bounding contours. The approach in [24] is first restricted to gait frontoparallel to the camera; however, they miss information about the arms in the contour data. This is an important aspect of gait. Fitting spatio-temporal surfaces to XYT cubes are computationally more difficult. The proposed method is very easy for computation, is view-invariant, and is very robust in obtaining sufficient individual characteristics of silhouette for gait recognition.

In [14], a spatio-temporal gait representation is generated by projecting the body silhouette along its columns and rows, then stacking these 1D projections over time to form a 2D pattern; they call this a frieze pattern. Then they used this pattern for gait alignment, but experiments for human identification were not reported. In [9], 2 different image features have been considered; the width of the outer contour of the binarized silhouette of the walking person and the entire binary silhouette itself. Then a hidden Markov model-based pattern recognition algorithm was used for gait recognition. In their experiments, the subjects were only walking laterally, with respect to the image plane. In [3], they first calculated the centroid of a silhouette as a reference origin. Then, they unwrap the outer contour counterclockwise to turn it into a distance signal. Next, eigenspace transformation, based on PCA, is applied to that distance signal. Euclidean distance-based supervised pattern classification techniques were finally performed for recognition. The basic idea of feature representation of a silhouette in [3] depends on calculating the centroid point of a silhouette. When silhouette data are noisy, e.g., small holes inside of objects, the centroid of a silhouette will most likely be faulty, and the distance signals may also be of different types for the same subject.

For gait cycle estimation, several vision methods have exploited this fact to compute the period of human gait from image features [3, 15, 16]. Gait period analysis serves to determine the frequency and phase of each observed sequence so as to align sequences before matching. In [15], width time signal of the bounding box of a moving silhouette derived from an image sequence is used to analyze the gait period. Either width time signal or height time signal is used because the silhouette height, as a function of time, plays an analogous role in periodicity [16]. In [3], the aspect ratio of the bounding box of a moving silhouette, as a function of time, is also used to determine the period of the gait. Different from these examples, this paper presents 4 projections-based variations on a moving silhouette as a function of time so as to enable them to cope effectively with both lateral and frontal views.

There are also earlier studies to detect a cyclic motion in the sequence [21-23]. In [21], cycles are detected using auto-correlation and Fourier transform techniques of the smoothed spatio-temporal curvature function of trajectories created by specific points on the objects as it performs cyclic motion. The detection of the specific points of humans, which they used, such as ankle, wrist, elbow, knee, hip, shoulder, and head, is a very difficult task, especially in real-time applications. In [22], they used similar techniques as in [21] to judge the degree of periodicity. They compute a reference curve (which is essentially a trajectory) by tracking the centroid of a moving region in several frames. They use the reference curve to align the frames, and then compute gray level signals at every pixel in the image frame. The gray level signals are then used to detect periodicity. Both algorithms are not fully automatic, and they assume the trajectory is given. In our proposed method, it is fully automatic, and there is no assumption of, the trajectory of motion objects. It is only assumed that the motion objects move along a straight path at free cadences. In [23], silhouettes of motion objects are first produced. The border of a silhouette is then extracted using morphological filtering. Then, the centroid of the silhouette border is determined and 1D signals are constructed by calculating the distances from the centroid to each border. Then, extremal points in the 1D distance signals are determined to represent special points of the human body, such as head, arms, and feet. A star skeletonization is then

produced using those points (centroid of the silhouette and special points). The uppermost skeleton segment represents the torso and measures the angle between this segment and vertical. The lower left segment also represents one of the legs and measures the angle between this segment and vertical. They finally used the variations of these 2 angles to distinguish between a running and walking person. The algorithm in [23] is not view-invariant. Any small noise in the silhouette can easily disturb the type of 1D signals by effecting the silhouette, because the centroid of the silhouette can be easily changed. The method we propose is view-invariant and any noise in the silhouette is not disturbing to the silhouette features used. The effects of any noise on the contour of the silhouette can also be reduced by smoothing the 1D signals produced from the 4 projections of the silhouette.

3. Feature Extraction

Before training and recognition, the silhouette shape of a human is first extracted by detecting and tracking persons in the surveillance area. Each image sequence, including a walking person's silhouette is converted into an associated temporal sequence of distance signals in the pre-processing stage.

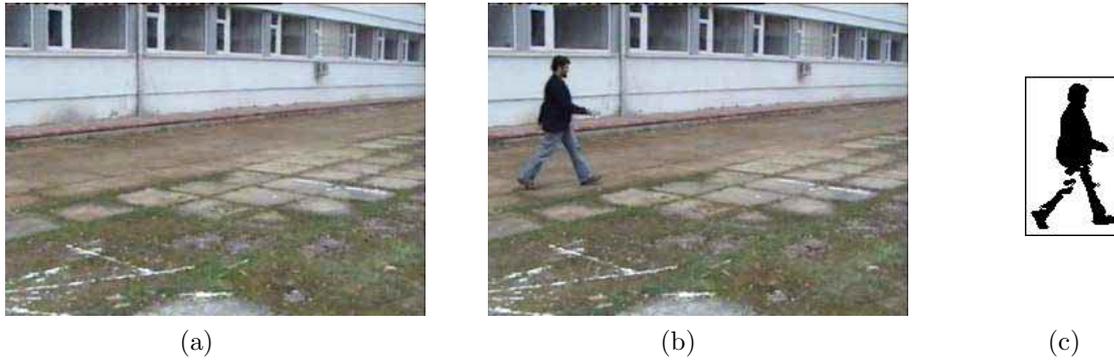


Figure 2. Example of moving silhouette extraction and a bounding box placed around it. (a) The background constructed, (b) an original image, (c) the extracted silhouette from (b), and bounding box.

3.1. Human Silhouette Extraction

Given a sequence of images obtained from a static camera, the algorithms developed before detect and track the moving person, then compute the corresponding sequence of motion regions in each frame. In order to detect motion objects, first, a background scene model is statistically learned using the redundancy of the pixel intensities in a training stage; even the background is not completely stationary. This redundancy information of each pixel is separately stored in a history map, which shows how the pixel intensity values changes over time. Then the highest ratio of the redundancy of the pixel intensity values in the history map in the training sequence is determined to create the initial background model of the scene. A background maintenance model is also proposed for preventing any kind of false information, such as illumination changes (the sun being blocked by clouds causing changes in brightness), or physical changes (person detection while he is getting out of or passing in front of a parked car). Based on the background modeling, candidate foreground regions are detected by using thresholding, noise cleaning, and morphological filtering; more details can be found in [7, 17]. Once a silhouette is generated, a bounding box is first placed around it. Silhouettes across a motion sequence are automatically aligned by scaling and cropping, based on the

bounding box. An example of silhouette extraction and a bounding box placed around it is shown in Figure 2.

3.2. Silhouette Representation

Silhouette representation is based on the projections of the silhouette, which are 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 4 different image features called distance vectors; top-, bottom-, left- and right-distance vectors. The distance vectors are the differences between the bounding box and the outer contour of the silhouette. An example silhouette and the distance vectors corresponding to 4 projections are shown in the middle of Figure 3. The distance vectors are separately represented by four 1D signals. The size of 1D signals for left- and right-distance vectors is the height of the bounding box. The values in the signals are data produced from differences between the bounding box and silhouette, which is the number of columns between the bounding box and silhouette at each row. The size of the 1D signals for top- and bottom-distance vectors is the width of the bounding box. The values of the 1D signals are the number of rows between the box and silhouette at each column.

Form 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 3, 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 column x of the 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 3, bottom-left.

The third pattern $F_L(y, t) = \sum_x bs(x, y, t)$ is then constructed by stacking left-projections and the last pattern $F_R(y, t) = \sum_{-x} bs(x, y, t)$ is finally constructed by stacking right-projections, as shown in Figure 3, top-right and bottom-right 2D patterns, respectively. For simplicity of notation, we write \sum_y , \sum_{-y} , \sum_x , and \sum_{-x} as shorthand for $\sum_{y=Topofthebox}^{Contourof silhouette}$, $\sum_{y=Bottomofthebox}^{Contourof silhouette}$, $\sum_{x=Leftsideofthebox}^{Contourof silhouette}$, and $\sum_{x=Rightsideofthebox}^{Contourof silhouette}$, respectively.

The variation of each component of each distance vector can be regarded as the 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 is in 2D patterns in Figure 3, the larger is the value of the distance vector in that position.

3.3. Gait Cycle Estimation

A gait cycle corresponds to one complete cycle from rest (standing) position to-right-foot-forward-to-rest-to-left-foot-forward-to-rest position [9]. So, gait cycle estimation is particularly important for gait identification. As discussed in section 2, several vision methods have exploited this fact to compute the period of human gait from image features [3, 15, 16]. This paper presents a view-invariant and more robust technique to estimate gait cycle.

The algorithm steps are shown in Figure 4 and as follows. The output of the detecting/tracking module gives a sequence of bounding boxes for every object [7]. The first step in the algorithm is to take the projections and to find the normalized correlation between consecutive frames. The elimination of the

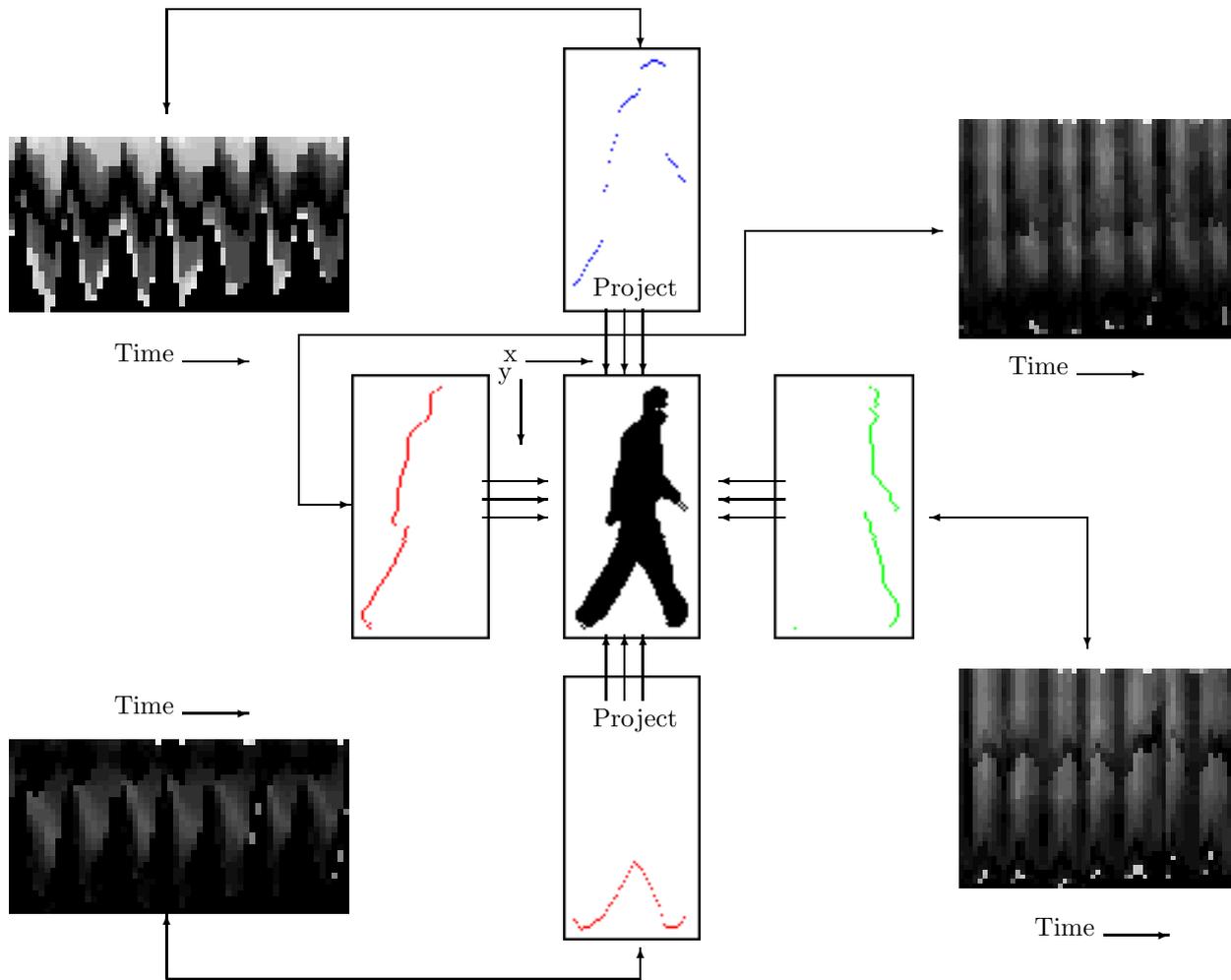


Figure 3. Silhouette representation. **(Middle)** Silhouette and 4 projections, **(Left)** temporal plot of the distance vectors for top and bottom projections, and **(Right)** temporal plot of the distance vectors for left and right projections.

influence of spatial scale and signal length are inherently achieved by selecting consecutive frames. In the cycle estimation, to quantify the signals of the correlation results, we may first remove their background component by subtracting their mean and dividing by their standard deviation, and then smoothing them with a symmetrical average filter. Further, we compute their autocorrelation to find peaks, as shown in Figure 4. In experimental studies, it has been reported that a more robust gait cycle estimation is achieved by using bottom- and top-projections, while the motion objects move laterally and obliquely, with respect to image plane. For objects that move frontally, gait cycle estimation is performed by using left- and right-projections. The direction of the motion of objects are determined by a motion tracking algorithm [7].

Some of the experimental results of the algorithm proposed and the aspect ratio signals, which were used in [3, 15, 16], are compared in Figures 5 and 6. The reason we present the experiments for running persons is to show the robustness of the gait cycle estimation algorithm presented here. The results of auto-correlation, as explained above, for a walking person in a frontal view and for a running person in an oblique view (Type 2) are shown in Figure 6 a and c, respectively. The results of auto-correlation on aspect

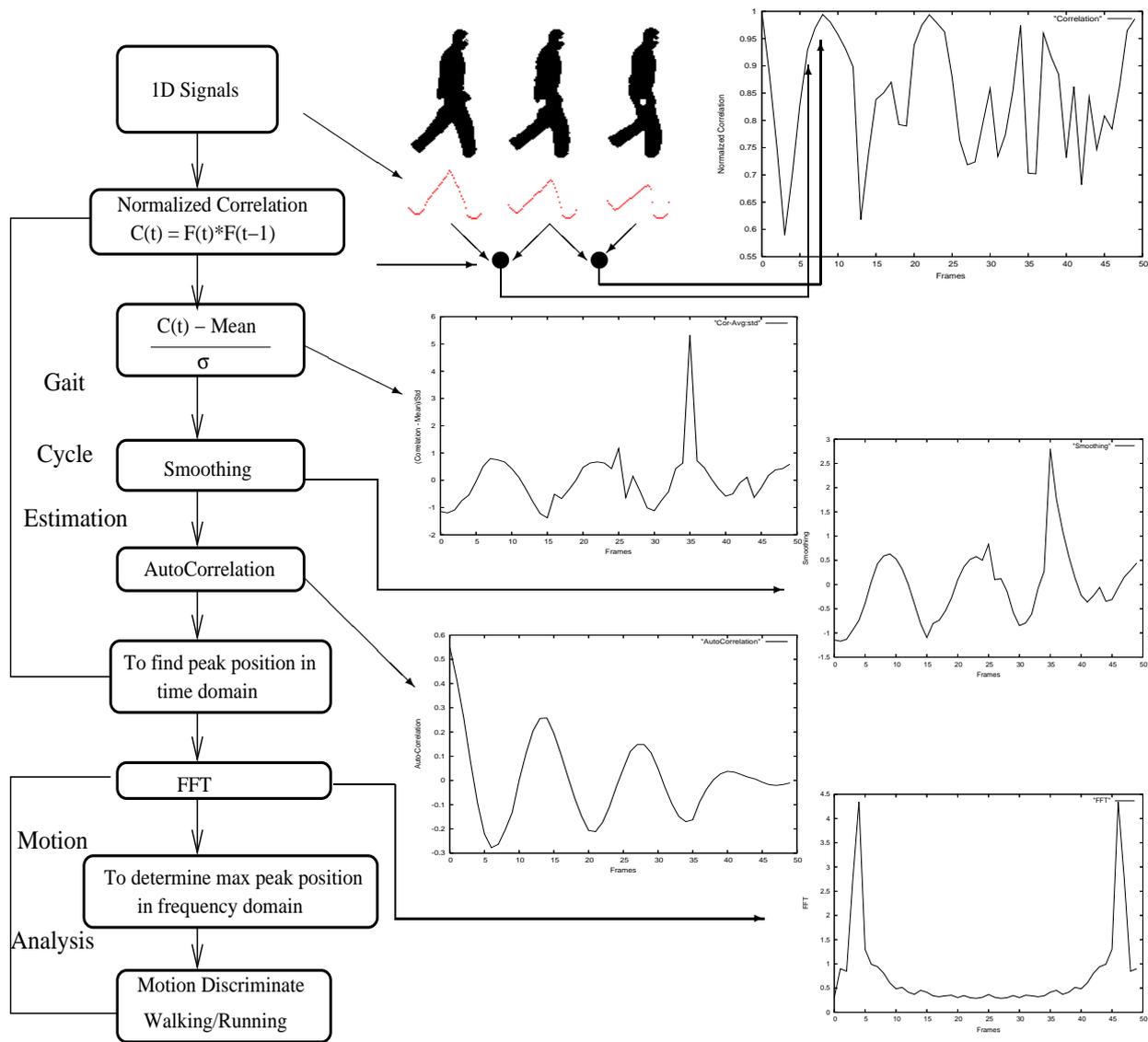


Figure 4. Gait cycle estimation and motion analysis.

ratio signals are also shown in Figure 6 b and d, respectively. As compared to the results in Figure 6 a and b, in the experiments with the gait sequence for frontal view, although the periodical characteristics of moving silhouettes are correctly detected by the left- and right-projections, based on the distance vectors (as plotted in Figure 6 a), but no periodical characteristics were achieved with the results of the aspect ratio signals, as shown in Figure 6 b. The gait cycle estimation presented here has achieved more robust experimental results than the aspect ratio-based gait cycle detection for frontal views. For lateral and oblique views, both algorithms can easily detect the gait cycles, as shown in Figure 6 c and d.

Some example silhouettes used and their corresponding projections in one gait cycle are shown in Figure 7, top. Normalized correlations for each projection during one gait cycle are separately plotted in Figure 7, middle. Results of auto-correlation after the background component of the correlation results are removed are shown in Figure 7, -bottom. In experimental studies, it has been reported that gait

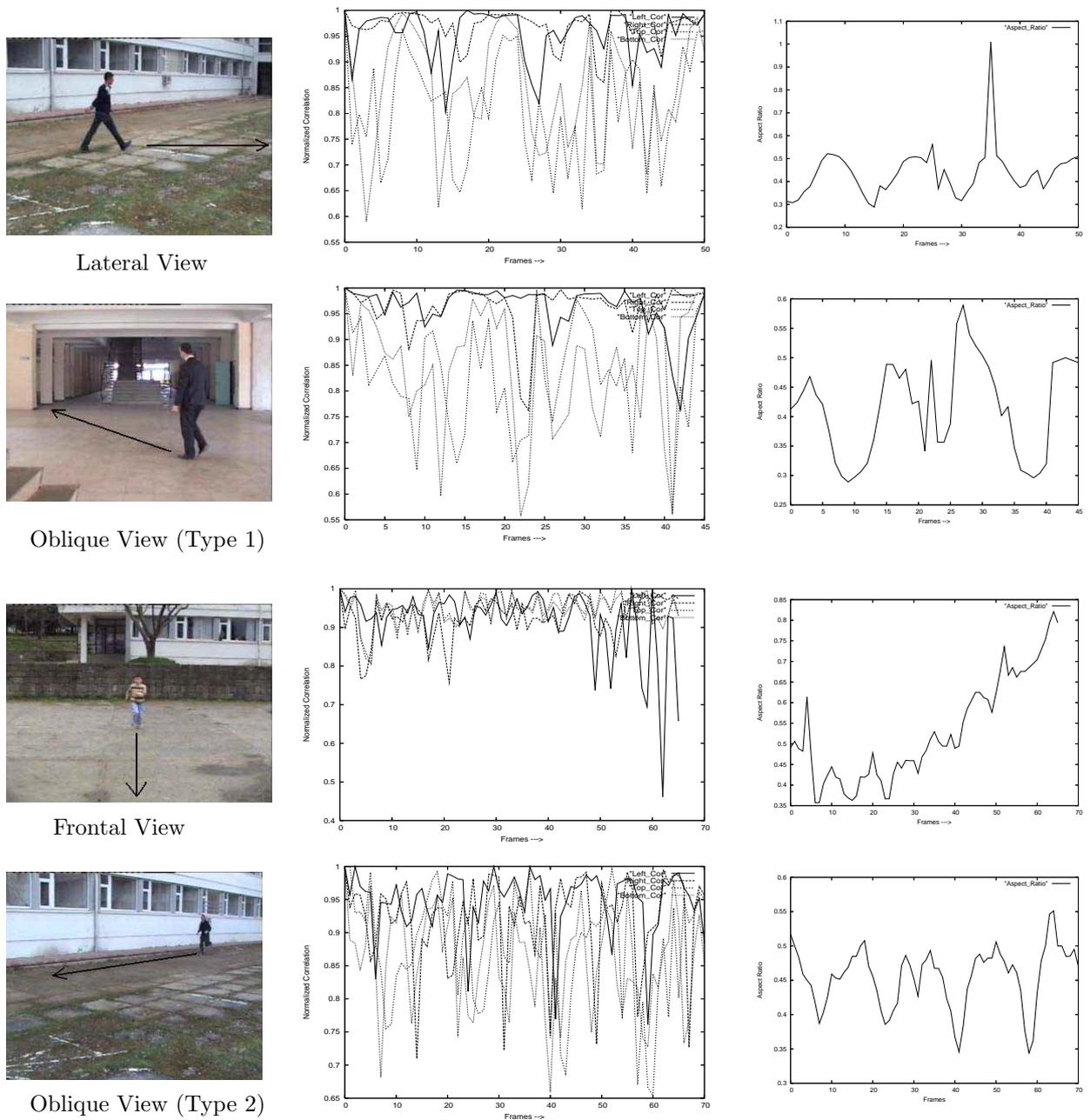


Figure 5. (Left) Some sample images in our database. (Middle) Normalized correlation and (Right) aspect ratio signals (i.e., width/height.) for walking and running persons.

cycle estimation is achieved using bottom- and top-projections when the motion objects move laterally and obliquely, with respect to image plane. For the other possibility, that is, when the objects move frontally, gait cycle is estimated based on left- and right-projections. The direction of the motion objects are determined by the object tracking algorithm [17].

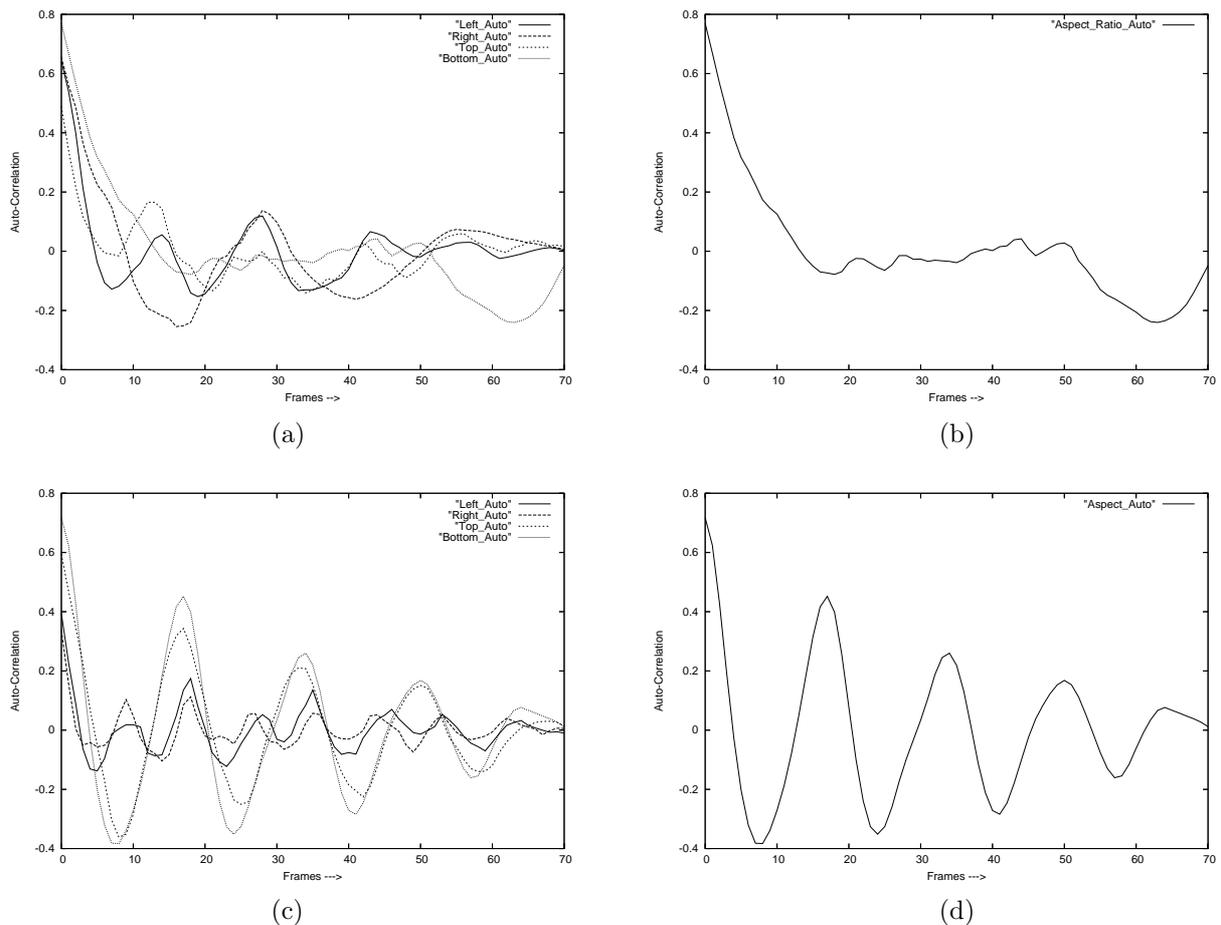


Figure 6. Comparing the results of gait cycle estimation. Auto-correlation results produced (a) by the proposed method and (b) by aspect ratio signals i.e., width/height for a walking person in the frontal view, respectively. Auto-correlation results produced (c) by the algorithm and (d) by the aspect ratio signal for a running person in the oblique view (Type 2), respectively.

3.3.1. Motion Discrimination

This section describes a low-level approach to distinguish walking and running actions from each other. The algorithm is used to estimate the frequency and phase of each observed gait sequence, which allow dynamic time warping to align sequences before matching to achieve gait-based human identification. To do that, it is also useful to move into the Fourier domain to distinguish walking and running actions. The autocorrelation signals are transformed into the frequency domain. In the frequency, the major cyclic components of the cyclic point can finally be extracted from the power spectrum of this processed signal. Some of the frames in the test sequences with humans in each image sequence perform walking or running movements in different directions and the test motion sequences were also obtained from lateral, frontal, and oblique (Type 1, Type 2) views. Some example frames are also shown in Figure 5.

To determine if an object exhibits periodicity, a 1-D power spectrum, $P(f)$, of auto correlation is estimated. Periodic motion will show up as peaks in this spectrum at the motion's fundamental frequencies

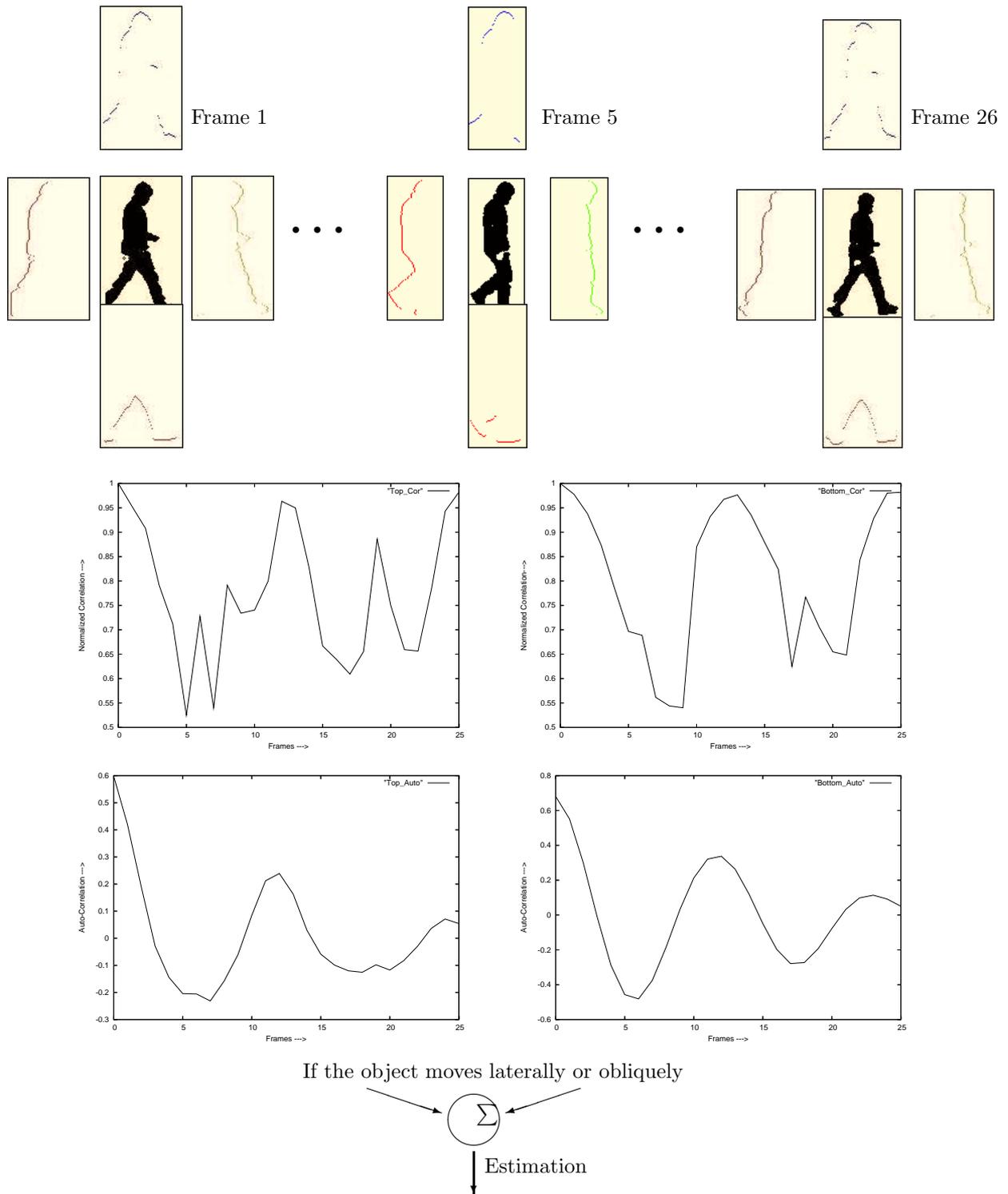
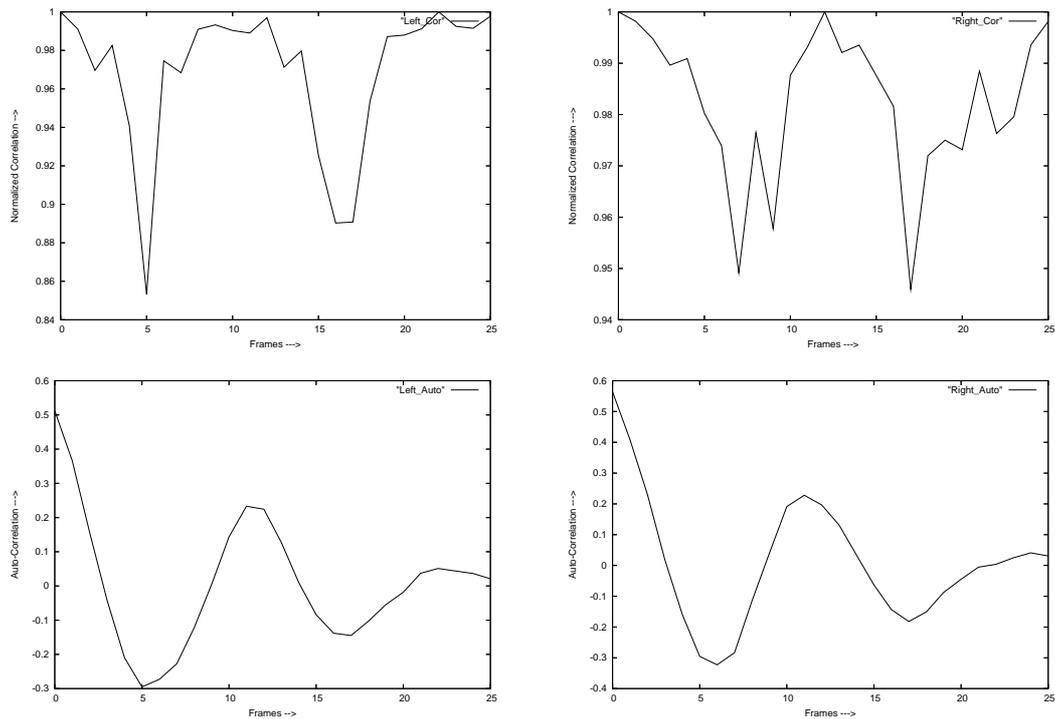


Figure 7. Main steps in the gait cycle estimation for one cycle includes 26 frames. Top: Silhouettes and their projections; Middle: Normalized correlation for top-bottom-left-right, respectively; Bottom: Auto-correlation. Deciding which projections of auto-correlation will be selected finally determines the gait cycle.



If the object moves frontally

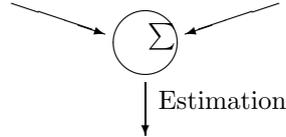


Figure 7. Countunied.

[25]. A peak at frequency f_i is significant if

$$P(f_i) = \mu_p + K * \sigma_p \tag{1}$$

where K is a threshold value (typically 0.7), μ_p is the mean of P, and σ_p is the standard deviation of P. The main motion frequency is estimated by determining the peak that has the largest impulse from the significant peaks (eq. 1) in the power spectrum. One cycle of the movements is extracted using the indicated location of the largest impulse. Smaller impulses may also be present (harmonics) at integer multiple of the fundamental.

3.4. Training

In the training, the algorithm first converts a two-dimensional silhouette shape into four 1D signals, called distance vectors, which will significantly reduce the subsequent computational cost. The distance vector sequence is, accordingly, used to approximately represent a temporal pattern of gait, as explained in section 3.2, and also illustrated in Figure 3. 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 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, which is similar to [3, 6] is illustrated as follows:

Given k classes for training and that each class represents a sequence of the distance vector signals of a single person, multiple sequences of each subject can be added for training, but in this work, a sequence including one gait cycle is used. Let $V_{i,j}^w$ be the j th distance vector signal in the class i 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 $N_t^w = N_1^w + N_2^w + \dots + N_k^w$, and the whole training set can be represented by $[V_{1,1}^w, V_{1,2}^w, \dots, V_{1,N_1}^w, V_{2,1}^w, \dots, V_{k,N_k}^w]$. The mean m_v^w and the global covariance matrix \sum^w of w projection training set can easily be obtained by

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

$$\sum^w = \frac{1}{N_t^w} \sum_{i=1}^k \sum_{j=1}^{N_i^w} (V_{i,j}^w - m_v^w)(V_{i,j}^w - m_v^w)^T \quad (3)$$

Here, each $V_{i,j}^w$ value represents the distance vectors for w projection (top-bottom-left-right). If the rank of matrix \sum is N , then the N nonzero eigenvalues of \sum , $\lambda_1, \lambda_2, \dots, \lambda_N$, and the associated eigenvectors e_1, e_2, \dots, e_N can be computed based on the theory of *singular value decomposition* [6]. The first few eigenvectors correspond to large changes in training patterns, and higher-order eigenvectors represent smaller changes [3]. As a result, for computing efficiency in practical applications, those small eigenvalues and their corresponding eigenvectors are ignored. This is done by using a threshold value. In the experiments, this threshold value chosen is 0.95 for obtaining steady results.

Then a transform matrix $T^w = [e_1^w, e_2^w, \dots, e_s^w]$ 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 the $s < N$ largest eigenvalues and their associated eigenvectors for each projection of the silhouette. This can be formalized as follows:

$$P_{i,j}^w = [e_1^w e_2^w \dots e_s^w]^T V_{i,j}^w. \quad (4)$$

Therefore, s values are usually much smaller than the original data dimension N . In other words, eigenspace analysis can reduce the dimensionality of input samples. Then the projection average A_i^w of each training sequence in the eigenspace is given by

$$A_i^w = \frac{1}{N_i^w} \sum_{j=1}^{N_i^w} P_{i,j}^w \quad (5)$$

4. Pattern Classification

For gait recognition, the similarity between reference (training) pattern and test samples (patterns) in the parametric eigenspace can be used.

4.1. Mahalanobis Distance

Mahalanobis distance measures has initially been selected for classification in this work. The accumulated distance between the associated centroids A^w (obtained in the process of testing) and B^w (obtained in the

process of training) can be easily computed by

$$d_S(A, B) = \sqrt{\left(\frac{A_1 - B_1}{s_1}\right)^2 + \dots + \left(\frac{A_p - B_p}{s_p}\right)^2} = \sqrt{(A - B)^t S^{-1} (A - B)} \quad (6)$$

where $S = \text{diag}(s_1^2, \dots, s_p^2)$. In the distance measure, the classification result for each projection is then accomplished by choosing the minimum of d . It has been observed in the experimental studies that when each distance vector is independently implemented from each other, recognition errors often occur, while similarity measuring gives positive for one projection, but not the other projection(s), and vice versa. So, as to increase the recognition performance, the classification results of each projection are finally fused using projection voting strategies.

4.2. Fusion

Two different strategies were developed. In **strategy 1**, each projection is separately treated. Then the strategy is to combine the distances of each projection at the end by assigning equal weight. The final similarity using strategy 1 is calculated as follows:

$$D_i = \sum_{j=1}^4 \alpha_j * d_{ji} \quad (7)$$

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). In conclusion, if any 2 of the distance similarity values in the 4 projections give maximum similarities for the same person, then the identification is determined as to be positive. Therefore, fusion strategy 1 has rapidly increased the recognition performance in the experiments.

In the experimental studies, it has been seen that some projections have given more robust results than others. For example, while a human moves in the lateral view, with respect to image plane, the back side of the human gives more individual characteristics of gait. The projection corresponding to that side can give more reliable results, and in such case, is called the dominant feature. As a result, **strategy 2** has also been developed to further increase recognition performance. In the strategy 2, if the dominant projection, or at least 2 projections of others, are positive for an individual, then the final identification decision is positive. The dominant feature in this work is automatically assigned by estimating the direction of motion objects under tracking [17]. In the next section, some of the dominant features determined experimentally for different view points, with respect to image plane, are also given.

5. Experiments and Results

The performance of the proposed algorithm is evaluated with CMU's MoBo database[20], NLPR gait database [3], and KTU database. The KTU database is established for testing the gait cycle estimation and gait recognition algorithm. The database mainly contains video sequences on different days in outdoor and indoor environments. A digital camera (Sony DCR-TRV355E) fixed on a tripod is used to capture the video sequences. For the test sequences, the detection and tracking algorithm used were provided by the work in [7]. The outputs of its method are the objects within the bounding boxes.

5.1. Gait Cycle Estimation

The test sequence includes 2 different actions, one contains 17 people (2 children and 15 adults) walking and running, the other sequence contains 22 persons (2 females, 20 males) walking only; the frame rate is 25 frames per sec, and the original resolution is 352 x 240. All subjects walk/run along a straight-line path at free cadences in different views laterally, frontally, obliquely, see Figure 5. In the experiments, while the subjects move frontally, with respect to the image plane, the left and right projection-based gait cycle estimations gave more reliable results, and an example of the results of the auto-correlation processing is shown in Figure 6 a. For the other view points, namely laterally and obliquely, although all projection-based gait cycle estimations gave reliable results, the cycle estimation based on bottom- and top-projections gave more robust results, as shown in Figure 6 c. These multi-options for gait cycle estimation have been automated by the tracking algorithm used [7].

5.2. Motion Discrimination

The gait cycle estimation algorithm was also used to discriminate human motions, such as walking and running, by re-implementing the cycle period in time, and by moving to the Fourier domain. In the time domain, the process is to estimate the period of the results of auto-correlation. Example silhouettes and their processing for motion analysis in lateral and frontal views are shown in Figures 8 and 9. As explained in section 3.3, bottom- and top-projections for lateral view, and right- and left-projections for frontal view are chosen to discriminate human motions, as well. This choosing procedure is automatically performed by the motion tracking algorithm [17]. In Figures 8 and 9, the results shown were obtained from bottom- and right-projections for lateral and frontal views, respectively. The differences of the results of auto-correlation for walking and running persons can easily be achieved to discriminate both motion actions.

Additionally, motion discrimination is also performed in the frequency domain. In this domain, the main frequency of motion is estimated by determining the peak that has the largest impulse from the significant peaks in the frequency power spectrum. Example experimental results in frequency are shown in Figures 8 and 9. It can be seen that the frequency is different for walking and running persons for lateral and frontal views. For the motion frequency determined by the largest impulse from the significant peaks, the average walking-running frequencies were found to be 3.571 (Hz) - 5.357 (Hz) for lateral, 3.26 (Hz) - 5.434 (Hz) for oblique (Type 1), 3.521 (Hz) - 5.633 (Hz) for oblique (Type 2), and 2.205 (Hz) - 2.941 (Hz) for frontal views, respectively. Threshold values in frequency were used to discriminate the walking and running actions. In this work, typical threshold values were taken as the mean of the average walking and running frequencies for each view. The values given were extracted from the video at a frame rate of 25 Hz. The motion discrimination algorithm presented has also used the cyclic information obtained from different projections of silhouettes. In other words, for frontal view, the cyclic knowledge produced from left- and right-projections has been implemented. For lateral and oblique views, bottom- and top-projections have been automatically selected to discriminate the motions, as explained before. Consequently, in the experimental results, the proposed method for motion discrimination correctly classified 94% of lateral views, 88% of oblique views (type1, and type 2), and 76% of frontal views, of the gaits.

5.3. Gait Recognition

The Viterbi algorithm was used to identify the test sequence, since it is efficient and can operate in the logarithmic domain using only additions. For every gait cycle, the processes rank order the probabilities

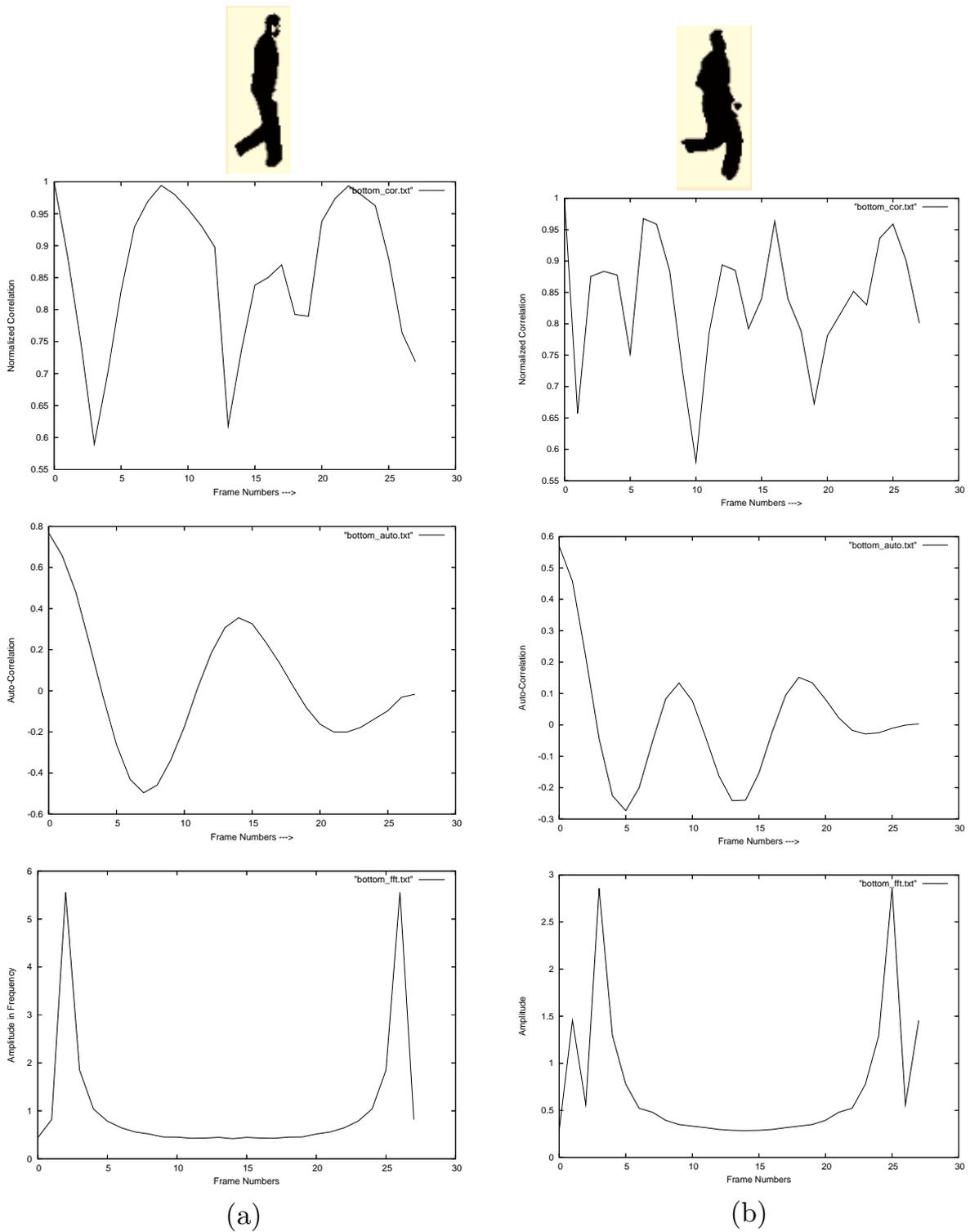


Figure 8. Motion discrimination: Example silhouettes in motion sequences, normalized correlation results, then auto-correlation, and frequency power spectrum (produced by FFT) of auto-correlation are shown, respectively. For lateral view: (a) walking; (b) running.

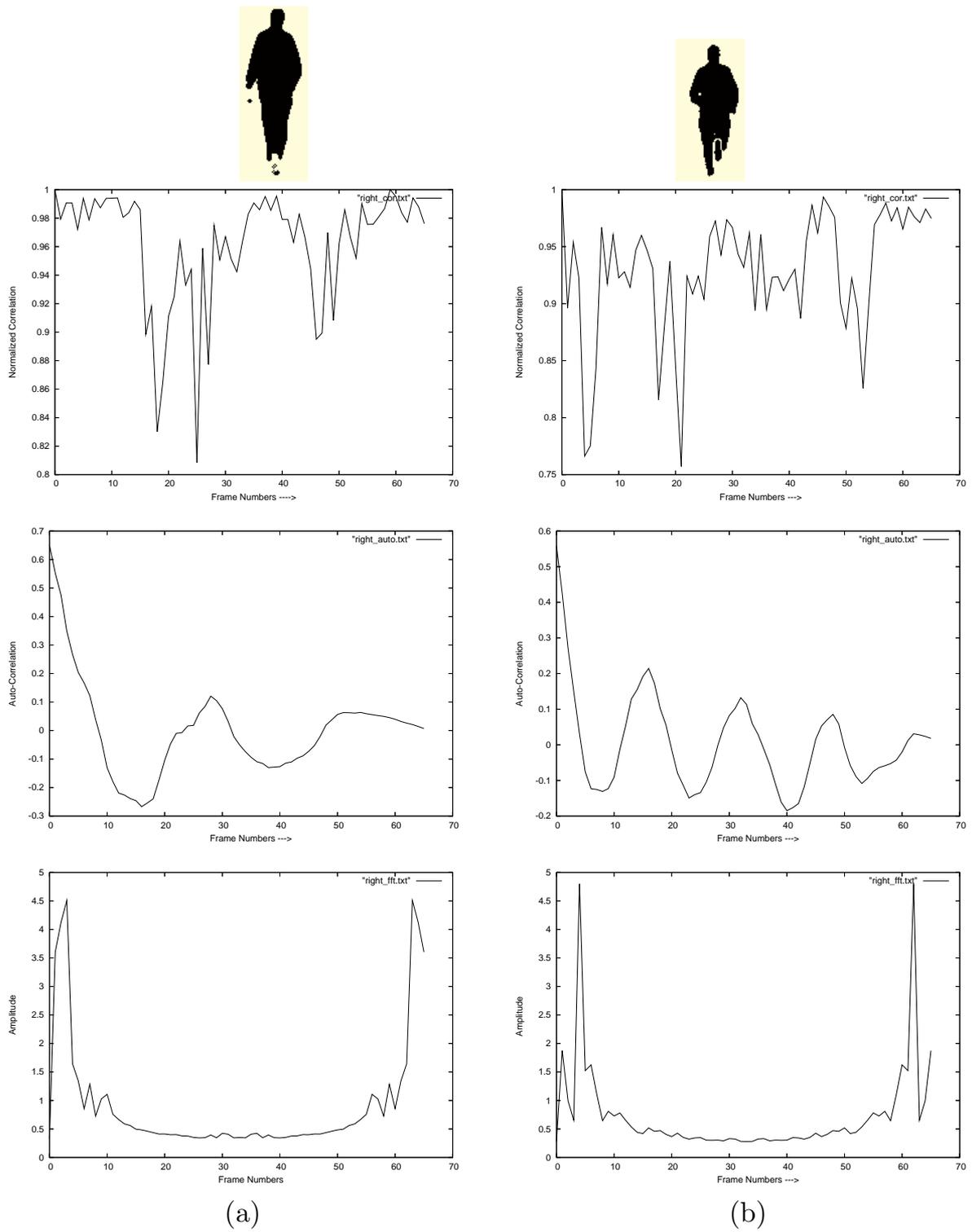


Figure 9. Motion discrimination: Example silhouettes in motion sequences, normalized correlation results, then auto-correlation, and frequency power spectrum (produced by FFT) of auto-correlation are shown, respectively. For lateral view: (a) walking; (b) running.

and the corresponding person indices in descending order. The algorithm then evaluates performance by letting each person in the database be the unknown p and plot the fraction of times that the right person is within the top r matches as a function of r . This curve is known as the cumulative match score (CMS) characteristics [18].

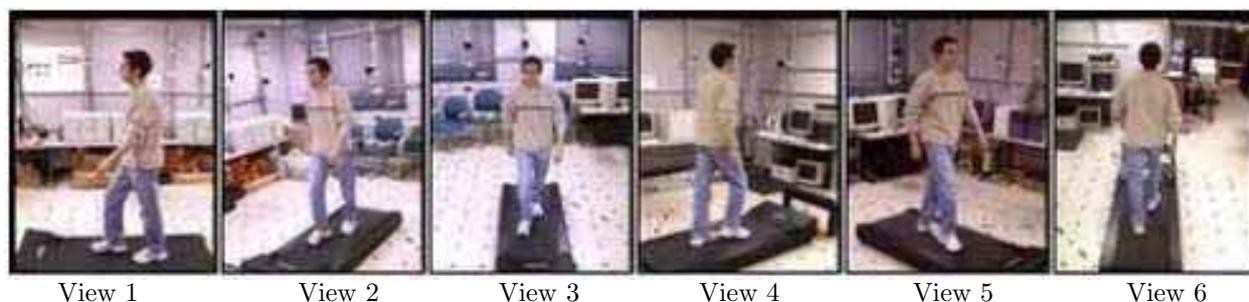


Figure 10. The six CMU Mobo database viewpoints.

5.3.1. CMU Database

The first experiments are obtained with the CMU database. This database has 25 subjects (23 males, 2 females) walking on a treadmill. Each subject is recorded performing 4 different types of walking: slow walk, fast walk, inclined walk, and slow walk holding a ball. There are about 8 cycles in each sequence, and each sequences is recorded at 30 frames per sec. It also contains 6 simultaneous motion sequences of the 25 subjects, as shown in Figure 10.

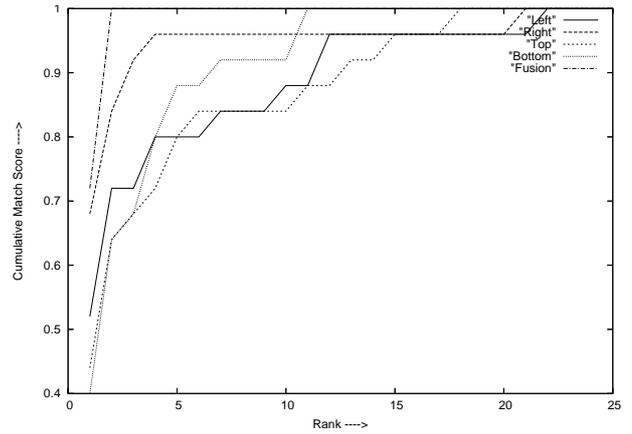
One of the cycles in each sequence was used for testing, others for training. First, I did the following experiments on this database: **1)** train with slow walk and test with slow walk; **2)** train with fast walk and test with fast walk; **3)** train with walking while carrying with a ball and test with walking while carrying a ball; **4)** train with slow walk and test with fast walk; **5)** train with slow walk and test with walking while carrying a ball; **6)** train with fast walk and test with slow walk; **7)** train with fast walk and test with walking while carrying a ball; **8)** train with walking while carrying a ball and test with slow walk; **9)** train with walking while carrying a ball and test with fast walk.

For case **1, 2, 3**, the experimental results obtained using the proposed method are given in Figure 11. Figure 11 shows the cumulative match score characteristics of Mahalanobis distance-based similarities of each the distance vectors and their fusion strategies. As illustrated in Figure 11, the fusion strategy developed in this work has increased the recognition performance and this is also one of the unique finding of this paper.

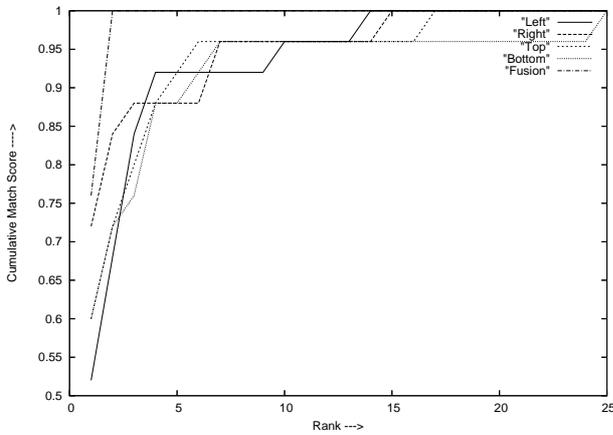
The results of all the cases obtained using the proposed method are also summarized **1-9** in Table 1. It can be seen that the right person in the top 2 matches 100% of the time for the cases in which testing and training sets correspond to the same walk style. When the strategy developed in the fusion as the 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 individuals, 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 3 matches 100% of the time for those cases, and the dominant projection strategy has also increased the recognition performance



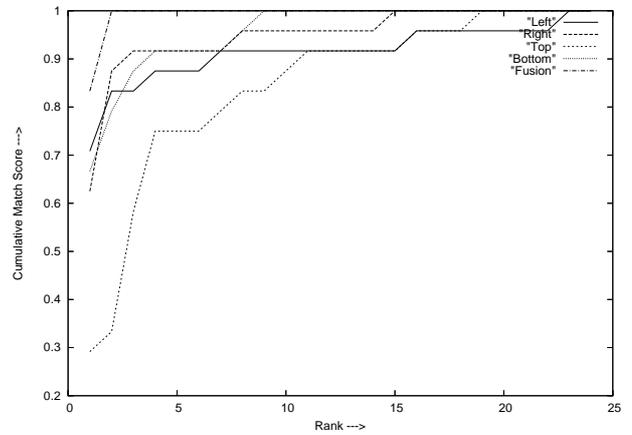
(a) Sample frame for walking and carrying a ball.



(b) Slow / Slow



(c) Fast / Fast



(b) Ball / Ball

Figure 11. Cumulative match characteristics in the MoBo database for viewpoint 1. Note: Testing and training sets correspond to the same walking styles, and they are represented by Test/Training.

Table 1. Classification performance of the CMU data set for viewpoint 1. Eight gait cycles were used; 7 cycles for training, 1 cycle for testing.

Test/Train	All projections: equal				Right projection dominant		
	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

for Ranks 1 and 2. For the case of training with walking while carrying a 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 shown in Table 1.

Table 2. Classification performance on the CMU data set for all views: Eight gait cycles were used; 7 cycles for training, 1 cycle for testing.

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

For the other view-points, the experimental results are also summarized for cases **1-4** in Table 2. When all the experimental results for the different view-points are considered, it can be seen that the right person in the top 2 matches 100% and in the top 4 matches 100% of the time for cases **1 and 2** and for cases **3 and 4**, respectively. It is also seen that when the dominant feature is used, gait recognition performance is also increased. Consequently, it is easy to see that the proposed method for gait recognition is view-invariant.

Some comparisons to the results with other reports in the literature for cases **1 and 6** are given in Table 3. The proposed method has recognition performance quite similar with the other reports for Rank 1. Fortunately, the algorithm presented has successfully given the right person in the top 2 matches (Rank 2) 100% of the time for the MoBo database, while other studies have not performed as well, as shown in Table 3. The numbers given for [9, 16] in Table 3 are as they appeared in the cited papers.

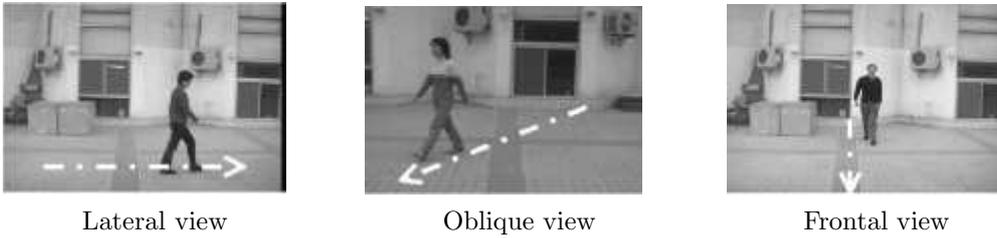
5.3.2. NLPR Database

The *NLPR* database [3] includes 20 subjects and 4 sequences for each viewing angle per subject, 2 sequences for one direction of walking, the other 2 sequences for the reverse direction of walking. For instance, when the subject is walking lateral to the camera, the direction of walking is from right to left for 2 of the 4 sequences, and from right to left for the remaining. All those gait sequences were captured twice (what we referred to as 2 experiments) on 2 different days, in an outdoor environment. All subjects walk along a straight-line path at free cadences in 3 different views with respect to the image plane, as shown in Figure

Table 3. Comparison of several algorithms on the MoBo dataset.

Algorithms	train/test	Rank 1(%)	Rank 2(%)	Rank 3 (%)	Rank 5 (%)
The method presented	slow/slow	84	100	100	100
Kale <i>et.al.</i> [9]	slow/slow	72	80	85	97
Collins <i>et.al.</i> [16]	slow/slow	86	100	100	100
The method presented	fast/slow	52	100	100	100
Kale <i>et.al.</i> [9]	fast/slow	56	62	75	82
Collins <i>et.al.</i> [16]	fast/slow	76	Not	given	92

12, where the white line with arrow represents one direction path and the other walking path is the reverse direction.

**Figure 12.** Some sample images in the NLPR gait database.

The following experiments were done with this database: **1:** train on one image sequence and test on the remainder, all sequences were produced from the first experiment; **2:** train on 2 sequences obtained from the first experiment and test on 2 sequences obtained from the second experiment. This is repeated for each viewing angle and for each direction of walking. The results for the experiments, along with the cumulative match scores in 3 viewing angles, are summarized in Table 4. When the experimental results are considered, the right person in the top 2 matches 100% of the time for lateral and frontal viewing angles, and in the top 3 matches 100% of the time for the oblique view.

Table 4. Performance on the NLPR data set for 3 views.

Direction	View	Training Set	Test Set	Rank1	Rank2	Rank 3
One Way Walking	Lateral	Experiment 1	Experiment 1	65	100	100
		Experiment 1	Experiment 2	55	100	100
	Frontal	Experiment 1	Experiment1	60	100	100
		Experiment 1	Experiment 2	35	100	100
	Oblique	Experiment 1	Experiment 1	40	90	100
		Experiment 1	Experiment 2	30	60	100
Reverse Way Walking	Lateral	Experiment 1	Experiment 1	60	100	100
		Experiment 1	Experiment 2	50	100	100
	Frontal	Experiment 1	Experiment 1	60	100	100
		Experiment 1	Experiment 2	40	100	100
	Oblique	Experiment 1	Experiment 1	45	100	100
		Experiment 1	Experiment 2	35	75	100

In the experiments on the NLPR database, the performance of the proposed algorithm was also

Table 5. Comparison of several algorithms on the NLPR database (Lateral View).

Methods	Rank 1(%)	Rank 2(%)	Rank 3 (%)	Rank 5 (%)	Rank 10 (%)
BenAbdelkader [15]	73	-	-	89	96
Collins [16]	71	-	-	79	88
Lee [11]	88	-	-	99	100
Phillips [26]	79	-	-	91	99
Wang [3]	75	-	-	98	100
The proposed method	65	100	100	100	100

compared with those of a few recent silhouette-based methods described in [3, 11, 16, 26]. To some extent, they reflect the latest and best work of these research groups in gait recognition. In [16], a method based on template matching of body silhouettes in key frames for human identification was established. The study in [11] described a moment-based representation of gait appearance for the purpose of human identification. A baseline algorithm was also proposed for human identification using the spatio-temporal correlation of silhouette images in [26]. The work in [3] computes the centroid of a silhouette's shape, and unwraps the outer contour to obtain a 1D distance signal, then applies PCA for human identification. These methods were implemented using the same silhouette data from the NLPR database with lateral view in [3], and the results of all algorithms are as taken from Tables in [3]. Table 5 lists the identification rates that have been reported by other algorithms and our algorithm. As a comparison, the proposed algorithm has given the right person in the top 2 matches 100% of the time for the NLPR database.

5.3.3. KTU Database

The database established for gait recognition includes 22 people (2 females, 20 males), and subjects are walking lateral to the camera, the directions of walking are from left to right, and from right to left. The database includes 2 sequences for each subject. One sequence includes 3 gait cycles 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; 15 subjects were walking outside, example frames are shown in Figure 5, and 7 subjects were walking inside, the inside platform is also shown in Figure 5. The results for the experiments, along with cumulative match scores in the lateral view are also summarized in Table 6. Walking from left to right and vice versa are separately tested to achieve initial experimental results. When the results of each projection-based distance vector are re-implemented using the dominant feature strategy, as explained in section 4, significant improvements in gait recognition was achieved. This is a robust implementation and is one of the unique findings of this paper.

Table 6. Performance on KTU data set: Three gait cycles were used, 2 cycles for training, 1 cycle for testing. (The abbreviation used :L \Rightarrow R : From Left to Right)

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

The silhouette quality in the NLPR and KTU databases was worse than the CMU database in terms of resolution and amount of noise. Hence, the recognition performance suffers most due to differences in

surface and background characteristics, and suffer the least due to loss of image quality. The increase in performance for Rank 1 is more likely to be achieved in the near future by performing some pre-processing on the distance vectors, such as noise reduction, and on the differences between successive vectors instead of the raw distance vectors, so as to be able to obtain more individual transits across silhouette motions. It is also possible to increase recognition performance by adding some additional physical features, such as pace, body height and build, and stride length, some of which have been independently used for personal identification in previous algorithms [3, 15, 19]. Nevertheless, the proposed method, which is not based on pre-processing the raw data of the distance vectors, has presented an encouraging performance for an initial study.

6. Conclusion

Gait refers to the style of walking of an individual. Recognition of a person from gait is a biometric of increasing interest. This paper described a multi-projection-based silhouette representation for individual recognition by gait. Unlike other gait representations, which consider only foreground pixels in a bounding box surrounding as silhouette and one aspect of gait, the proposed method represents human motion as a multi-sequence of templates for each individual and considers all background pixels in the bounding box. Unlike other gait cycle estimation algorithms, which analyze the variation of the bounding box data, the periodicity of gait is produced by analyzing silhouette data itself. The proposed algorithm has robustly estimated the periodicities in gait sequences obtained from all 3 views with respect to the image plane (lateral, oblique, and frontal). The approach for both gait cycle estimation and gait recognition is also fully automatic for real-time applications. Experimental results show that the proposed algorithm is an effective and efficient gait representation. The proposed recognition approach also achieves highly competitive performance with respect to the major published recognition approaches.

7. Acknowledgment

This paper was written with the partial supported by of The Research Foundation of Karadeniz Technical University (Grant No. KTU-2002.112.009.1). The author would like to thank Eyup Gedikli for producing silhouettes with the developed algorithm, and Mubarak Shah of the University of Central Florida, USA, for his helpful discussions on the technical sections of the paper.

References

- [1] Zongyi Liu, Sudeep Sarkar, *Effect of Silhouette Quality on Hard Problems in Gait Recognition*. IEEE Transactions on Systems, Man, and Cybernetics- Part B: Cybernetics, Vol.35, No. 2, April 2005.
- [2] M. S. Nixon, J. N. Carter, *Advances in Automatic Gait Recognition*, Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, 2004.
- [3] L. Wang, T. Tan, H. Ning, W. Hu, *Silhouette Analysis-Based Gait Recognition for Human Identification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No. 12, December, 2003.
- [4] C. BenAbdelkader, R. G. Cutler, L. S. Davis, *Gait Recognition Using Image Self-Similarity*. EURASIP Journal of Applied Signal Processing, pp. 1-14, April, 2004.

- [5] G. V. Veres, L. Gordon, J.N. Carter, M. S. Nixon, *What image information is important in silhouette-based gait recognition?* Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2004.
- [6] P. Huang, C. Harris, M. Nixon, *Human Gait Recognition in Canonical Space Using Temporal Templates*, IEE Proc. Vision Image and Signal Processing Conf., 1999.
- [7] M. Ekinici, E. Gedikli, *Background Estimation Based People Detection and Tracking for Video Surveillance*. Springer-Verlag Lecture Notes in Computer Science LNCS 2869, pp.421-429, November, 2003.
- [8] S. Sarkar, et.al., *The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No. 2, February, 2005.
- [9] A. Kale, A. Sundaresan, A.N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, R. Chellapa. *Identification of Humans Using Gait* IEEE Transactions on Image Processing, Vol.13, No.9, September 2004.
- [10] J. Cutting, L. Kozlowski, *Recognizing Friends by Their Walk: Gait Perception without Familiarity cues*. Bull. Psychonom. Soc. Vol.9, pp.353-356, 1977.
- [11] L. Lee, W. Grimson, *Gait Analysis for Recognition and Classification*, Proc. IEEE, Int. Conference on Automatic Face and Gesture Recognition, pp. 155-162, 2002.
- [12] M. Ekinici, E. Gedikli, *A Novel Approach on Silhouette Based Human Motion Analysis for Gait Recognition*. Springer-Verlag Lecture Notes in Computer Science, LNCS Vol. 3804, pp. 219-226, December, 2005.
- [13] M. Ekinici, E. Gedikli, *Gait Recognition Using View Distance Vectors*. Springer-Verlag Lecture Notes in Artificial Intelligence, LNAI Vol. 3801, pp.973-978, December, 2005.
- [14] Yanxi Liu, R. T. Collins, T. Tsin, *Gait Sequence Analysis using Frieze Patterns*, Proc. of European Conf. on Computer Vision, 2002.
- [15] C. BenAbdelkader, R. Cutler, L. Davis, *Stride and Cadence as a Biometric in Automatic Person Identification and Verification*, Proc. Int. Conf. Aut. Face and Gesture Recog.,2002.
- [16] 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.
- [17] M. Ekinici, E. Gedikli *Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance*, Turkish Journal of Electrical Engineering and Computer Sciences, Vol.13, No. 2, pp.199-230, 2005.
- [18] J. Phillips *et.al*, *The FERET Evaluation Methodology for Face recognition Algorithm*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, No.10, October 2000.
- [19] A. Bobick, A. Jhonson, *Gait Recognition Using Static, Activity-Specific Parameters*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2001.
- [20] 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.
- [21] P. Tsai, M. Shah, K. Keiter, T. Kasparis, *Cyclic Motion Detection for Motion Based Recognition*, Pattern Recognition, Vol. 27, No.12, 1994.
- [22] R. Polana, R. Nelson, *Detecting Activities*, Proc. of Computer Vision and Pattern Recog., 26:595-610, 1993.

- [23] H. Fujiyoshi, A. J. Lipton, T. Kanade.: Real-Time Human Motion Analysis by Image Skeletonization. IEICE Trans. Inf.& SYST.,Vol.E87-D, No.1, January 2004.
- [24] S. A. Niyogi, E. H. Adelson, *Analyzing and Recognizing Walking Figures in XYT*, in IEEE Conf. on Computer Vision and Pattern Recognition, June 1994.
- [25] R. Cutler, L. S. Davis, *Robust real-time periodic motion detection, analysis and applications* IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 22, 2000.
- [26] P. Phillips, et.al., *Baseline Results for Challenge Problem of Human ID using Gait Analysis*, Por. Int. Conf. Automatic Face and Gesture Recognition, 2002.