A Novel Approach on Silhouette Based Human Motion Analysis for Gait Recognition

Murat Ekinci and Eyup Gedikli

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

Abstract. This paper[1] presents a novel view independent approach on silhouette based human motion analysis for gait recognition applications. Spatio-temporal 1-D signals based on the differences between the outer of binarized silhouette of a motion object and a bounding box placed around silhouette are chosen as the basic image features called the distance vectors. The distance vectors are extracted using four view directions to silhouette. Gait cycle estimation and motion analysis are then performed by using normalized correlation on the distance vectors. Initial experiments for human identification are finally presented. Experimental results on the different test image sequences demonstrate that the proposed algorithm has an encouraging performance with relatively robust, low computational cost, and recognition rate for gait-based human identification.

1 Introduction

The combination of human motion analysis and gait recognition based human identification, as biometrics, in surveillance systems has recently gained wider interest in the research studies [1][3][6]. There has also been considerable interest in the area of human motion classification [12], tracking and analysis [4] in recent years. Those are required as initial steps in gait recognition algorithms for human identification applications [1][3]. The main purpose and contributions of this paper are summarized as follows;

- We attempt 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], but the idea presented here implicitly more capture both structural (appearances) and transitional (dynamics) characteristics of gait.
- Instead of width/length time signal of bounding box of moving silhouette usually used in existing gait period analysis [10][11][3], here we analyze four distance vectors extracted directly from differences between silhouette and the bounding box, and further convert them into associated four 1D signals.

[1] This work is suported by KTU (Grant No: KTU-2002.112.009.1).

G. Bebis et al. (Eds.): ISVC 2005, LNCS 3804, pp. 219–226, 2005.

– The proposed method also presents the motion cues which can be used to determine human activities such as walking or running. Unlike other methods, this does not require on a geometrical shape models [9], not sensitivity to small noisy on silhouette data [4]. The proposed algorithm is not pixel-based implementation, and the idea can be applicable to real scenes in which objects are small and data is noisy.

The method presented is integrated into a low-cost PC based real-time view independent visual surveillance system for gait cycle estimation, human motion analysis their activities in monochromatic video, and then adapted into initial studies on gait recognition for human identification. The novel approach is basically to produce view directions based 1-D distance vectors represent the distances between silhouette and the bounding box. Thus, four 1-D signals are extracted for each view directions, they are top-, bottom-, left-, and right-views. Then correlation-based a similarity function is executed to estimate gait cycle of moving silhouette and to analysis human motions. The key idea in this work is that simple, view independent, fast extraction of the broad internal motion features of an object can be employed to analyze its motion and meanwhile for gait-based human identification. Finally, gait recognition results are also presented to show applicable the proposed method for human identification.

2 Spatio-temporal Human Motion Representation

Spatio-temporal human motion representation is based on the view directions 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: top-, bottom-, left- and right-view based distance vectors, as shown in figure 1. 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-view distance vector of silhouette image *bs(t)*, as shown in figure 1.a. 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 boundaries in that columns x of silhouette image *bs(t)*. The result is a 2D pattern, formed by taking the differences from the top view direction together to form a spatio-temporal pattern. A second pattern which represents the bottom-view direction $F_B(x,t) = \sum_y bs(x,y,t)$ can be constructed by counting the number of the row pixels between the bottom of the bounding box and its silhouette, as shown in figure 1.a. The third pattern $F_L(y,t) = \sum_x bs(x,y,t)$ is then constructed by taking the row differences (the number of the column pixels) from the left side of the bounding box to silhouette. The last pattern $F_R(y,t) = \sum_x bs(x,y,t)$ is also finally constructed by taking the row differences from the right side of the box to silhouette (figure 1.c).

From the temporal distance plots, it is clear that the view distance vector is roughly periodic and gives the extent of movement of the outer contours on the view direction of silhouette. The brighter a pixel in figure 1.b and 1.d, the larger value is the value of the view direction vector in that position. In this study, silhouette extraction is achieved by simple background subtraction using

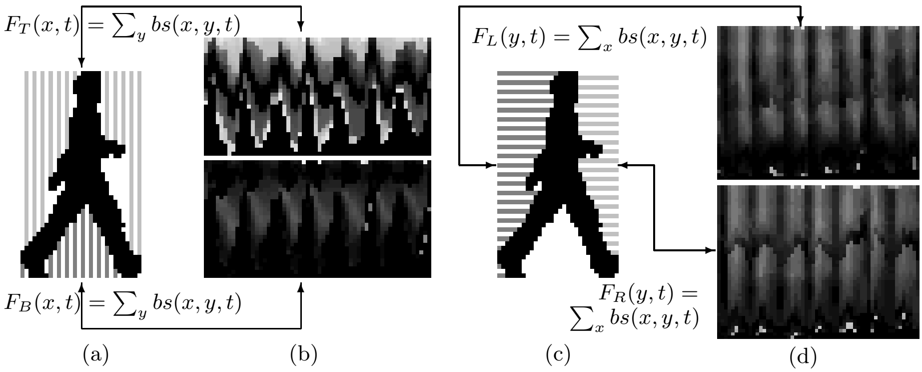

Fig. 1. Spatio-temporal motion representations. (a) Top-, bottom- (c) Left-, right-views to silhouette, (b) and (d) temporal plot of the distance vectors.

a dynamic background frame estimated and updated in time, more details could not given here because of page limitation of the paper, but it can be found in [2]. Then a 3x3 median filter operator is applied to the resulting images to suppress spurious pixel values. Once a silhouette generated, a bounding box is placed around silhouette. Silhouette across a motion sequence are automatically aligned by scaling and cropping based on the bounding box.

2.1 Features Derived from View Distance Vectors

The output of the detecting and tracking module gives a sequence of bounding boxes for every object [2]. Reference signals $R_v(x)$ and $R_v(y)$ for v view distances vectors at location column and row to silhouette are first obtained by assigning the distance signals produced from previous frame, respectively. In other words, the processes, $R_v(x,t) = F_v(x,t-1)$ and $R_v(y,t) = F_v(y,t-1)$, are executed. Then the same manner is sequentially repeated for the following frames to update the reference signals. In order to use in gait cycle estimation and motion discrimination, normalized correlation is performed to obtain maximum similarities between two view distance vectors produced from sequential two frames; for instance $F_T(x,t)$ in the current frame and $R_T(x,t)$ from the previous frame for top view. This is repeated for each view distance vectors.

3 Motion Analysis

We present a low-level approach to distinguish walking and running actions and to estimate the frequency and phase of each observed gait sequence, allowing us to perform dynamic time warping to align sequences before matching to achieve a gait based human identification. First, normalized correlation processes are executed as explained in previous section, and some of the experimental results are also shown in figure 2. For an input sequence (Fig. 2.a), once the person has been tracked for a certain number of frames, its spatio temporal gait parameters such as the normalized correlation based variations of the moving silhouette can be estimated.

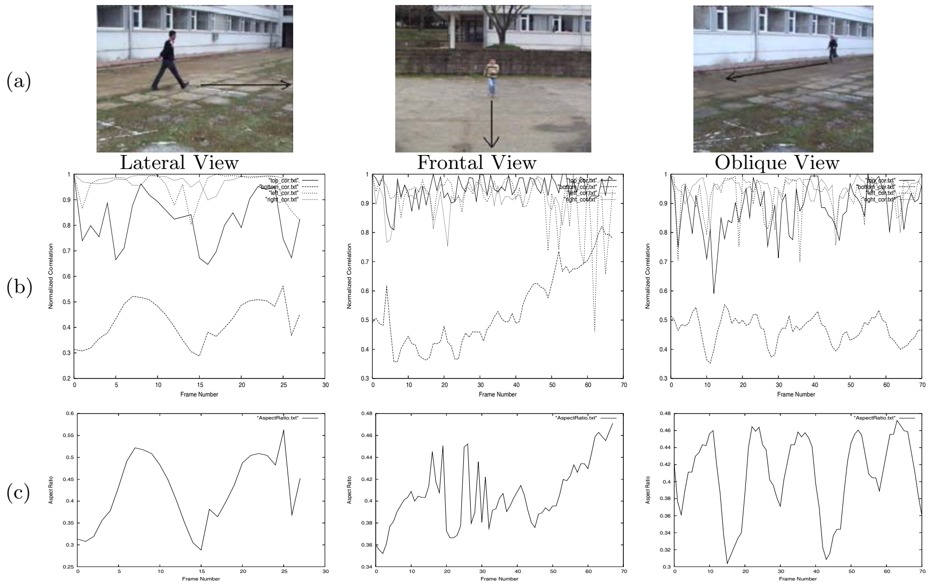

Fig. 2. (a) Example images used, (b) Normalized correlation results for walking and running, (c) Aspect ratio signals (i.e width/height.) of the bounding box

The normalized correlation results on four view distance vectors in the image sequence includes walking and running person's actions grabbed on different views are acquired as shown in figure 2.b. In that, the reference signals are automatically updated by copying the view distance vectors produced from previous frame in the sequence. To be able to produce optimum similarities on the distance vectors, the reference signals have naturally been normalized by selecting from previous frame in order to eliminate the influence of spatial scale and signal lengths.

The last line in figure 2 (i.e. Fig. 2.c) shows the aspect ratio signals (i.e. width/height in the bounding box) of the moving silhouettes. From the correlation plot, it is noted that view distance vectors change with time as the person transits through a period of view independent action, there is a high degree of correlation among the distance vectors across frames. In the experimental studies in the image sequence includes walking and running persons in lateral view, the normalized correlation results obtained from the all view distance vectors presented in this paper have exhibited the periodical signals. For oblique views, the top- and bottom-views distance vectors based normalized correlation processes have given more robust results than the left- and right-views distance vectors. The characteristics in their results on the test image sequences have given the similar signal characteristics with the aspect ratio signals for both lateral and oblique views, as shown in figure 2. For frontal view, the aspect ratio signals have not given a periodical signal on both actions (walking and running). But the view distance vectors-based correlation results have successfully exhibited the periodical signal characteristics.

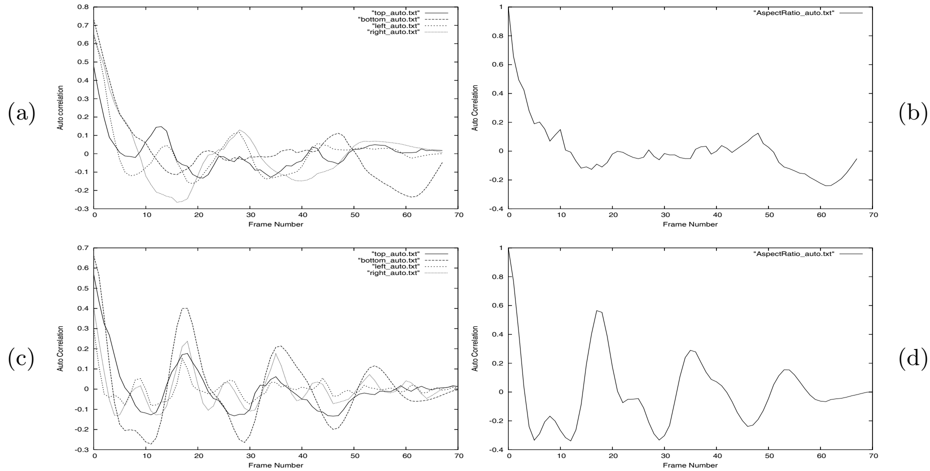

Fig. 3. Gait cyclic Estimation: Signals after removing the background then auto-correlation. (a) View distance based (b) Aspect ratio based auto-correlation results for walking person in frontal view, (c) View distance based (d) aspect ratio based auto-correlation results for running person in oblique view.

3.1 Gait Cycle Estimation

Human gait is a repetitive phenomenon, the appearance of a walking/running person in a video is itself periodic. Several vision methods have exploited this fact to compute the period of human gait from image features [10][11][3]. In [10][11], width time signal or height time signals of the bounding box of moving silhouette derived from an image sequence are used to analyze gait period. In [3], the aspect ratio of the bounding box of moving silhouette as a function of time is also used to determine the period of the gait. Different from them, here this paper presents four view distance vectors based variations on the moving silhouette as a function of time so as to enable them to cope effectively with both lateral view and frontal view.

Figure 2 displays a clear cyclical nature in the correlation results obtained on the view direction-based distance vectors. To quantify these signals, we may first remove their background component by subtracting their mean and dividing by their standard deviation, and then smooth them with a symmetric average filter. We finally compute their autocorrelation to find peaks, as shown in figure 3. For frontal view, although the periodical characteristics of moving silhouettes are correctly detected by left- and right-view distance vectors based gait cycle estimation (as plotted in figure 3.a), there has not been able to achieved any periodical characteristics on the results of the aspect ratio signals on the bounding box, as shown in figure 3.b. The proposed algorithm for gait cycle estimation has achieved more robust experimental results than the aspect ratio for frontal view. For oblique, as plotted in figure 3c-d, and lateral views, both the proposed method and the aspect ratio can easily detect the gait cycles.

3.2 Motion Discrimination

To determine if an object exhibits periodicity, an 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 [5]. 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. In order to distinguish walking and running, the main motion frequency is estimated by determining the peak which has 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 multiplies of the fundamental.

Another point of interest to distinguish two actions from each other is that the variance of running frequencies is greater than that of walking frequencies. For each view distance vectors, the average variance of the amplitude in the power spectrum was determined on the experimental video sequences produced from different views (Lateral, frontal, oblique). Then the average variance values for each view distance vectors have also been used as threshold data. The threshold values for the variance of the motion, the running frequencies can be clearly distinguished from the that of walking. Those values only correctly classify 88% of the gaits (averaging of three different motion way). For the motion frequency determined by the largest impulse from the significant peaks, the average walking and running frequencies were found to be 3.571 (Hz)- 5.357 (Hz) for lateral, 3.521 (Hz)- 5.633 (Hz) for oblique, and 2.205 (Hz)- 2.941 (Hz) for frontal views, respectively. The average threshold values in frequency were used to discriminate the walking and running actions. At this work, typical threshold values were taken the mean of the averaging walking and running frequencies for each view. The values given were extracted from the video at a frame rate of 25 Hz. Then, four decision values based implementation correctly classify 94% for lateral view, 88% for oblique view, 76% for frontal view, of the gaits.

Finally, both threshold values on the variance of motion frequencies and largest impulses in the power spectrum were implemented to have more robust decision for motion analysis, then averaging correctly classify 94% of the human motion analysis even the silhouette data is produced from in noisy environments. At the experiments for motion discrimination, the database has 17 person (2 child, 15 adults), and their action includes walking and running in outdoor environment with three different way. Example frames are shown in figure 2.

3.3 Gait-Based Human Identification

We convert a two-dimensional silhouette shape into four one-dimensional distance vectors. The distance vector sequence is accordingly used to approximately represent a temporal pattern of gait. This process of original feature extraction is illustrated in Figure 1. To eliminate the influence of spatial scale and signal length of the distance vectors, we scale these distance vector signals with respect to magnitude and size through the sizes of the bounding boxes placed around silhouette. 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. A normalized Euclidean distance based pattern classification technique is final performed in the lower-dimensional eigenspace for recognition. The training process used is similar to the studies in [3].

4 Experimental Results and Conclusion

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. The algorithm presented here has been tried on a database includes 17 people walking and running for motion discrimination, and frame rate was 25 fps and the original resolution is 352x240. Test results given in section 3 encourage to implement this kind of parameters for human motion analysis for real time video surveillance applications, as long as human performs the same actions during the test sequence.

At the initial experiments on gait recognition, a gait database is established for our experiments. The database has 22 people, 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 cadence in lateral view with respect to the image plane, and 15 subjects were walking outside, an example frame is shown in figure 2.a-left, 7 subjects were walking inside, the inside platform is also shown in figure 2.a, but the subjects were walking in lateral view.

The initial results obtained using the proposed method are given in figure 4. Figure 4 shows the cumulative match scores for ranks up to 22, where figure

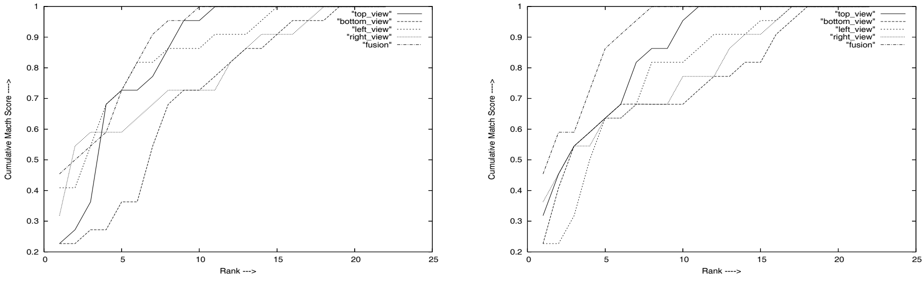

Fig. 4. Cumulative match score characteristics in the database, the subjects were walking (Left) from left to right, (Right) from right to left

uses the normalized Ecludiean distance similarity measures for each distance vectors and their fusion. Two cycles were used for training, and next one cycle was used for testing. Walking from left to right and the other direction are separately tested to obtain initial experimental results. The more details on the cumulative match score can be found in [13]. While only using alone each distance vectors without making a relationship between each others, we can see from experimental observations that recognition errors often happen when the two smallest values of the similarity function are very close for one view but not other view(s). Therefore, to have more robust classification, the classification processes of the distance vectors are finally re-implemented by fusing Euclidean distance measures. Then an increasing performance for the recognition has also been recorded at the initial experimental study, as shown in figure 4.

The novelty in the presented algorithm lies in its simplicity, the efficiency of the implementation, the usefulness in real-time applications, the effectiveness in view independent applications, the robustness to some factors such as motion object size and regular noise effects, and new approach includes multi-view and multi-purpose aspects to silhouette for silhouette based human identification.

References

1. M. S. Nixon, J. N. Carter, *Advances in Automatic Gait Recognition*, Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2004.
2. M. Ekinci, E. Gedikli, *Background Estimation Based People Detection and Tracking for Video Surveillance.* Springer LNCS 2869, ISCIS, 18*th* Int. Symp., Nov., 2003.
3. L. Wang, T. Tan, H. Ning, W. Hu, *Silhouette Analysis-Based Gait Recognition for Human Identification,* IEEE Trans. on PAMI, Vol.25, No. 12, December, 2003.
4. H. Fujiyoshi, A. J. Lipton, T. Kanade.: Real-Time Human Motion Analysis by Image Skeletonization. IEICE Trans. Inf.& SYST., Vol.E87-D, No.1, 2004.
5. R. Cutler, L. S. Davis, *Robust real-time periodic motion detection, analysis and applications* IEEE Trans. Pattern Analysis Machine Intelligence, Vol. 22, 2000.
6. S. Sarkar, *et. al*, *The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis.* IEEE Trans. on PAMI, Vol.27, No. 2, February 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. J. W. Davis, S. R. Taylor, *Analysis and Recognition of Walking Movements.* Proc., of int. Conference on Pattern Recognition, ICPR 2005.
10. C.BenAbdelkader,R. Cutler,L.Davis, *Stride and Cadence as a Biometric in Automatic Person Identification and Verification,* Proc. of Aut. Face Gest. Rec, 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. O. Javed, S. Ali, M. Shah, *Online Detection and Classification of Moving Objects Using Progressively Improving Detectors*, IEEE, Proc. of Int. CVPR 2005
13. J. Phillips *et.al*, *The FERET Evaluation Methodology for Face recognition Algorithm,* IEEE Trans. Pattern Analysis and Machine Intel., vol.22, no.10, Oct. 2000.