Real-Time Human/Vehicle Classification and Analysis by View Direction Distances

Murat Ekinci

Computer Vision Lab.
Department of Computer Engineering
Karadeniz Technical University
Trabzon 61080, Turkiye,
ekinci@ktu.edu.tr

Abstract. This paper presents a novel view independent approach is described for real-time human/vehicle classification and motion analysis in real visual surveillance scene. Spatio-temporal 1-D signals based on the differences between the outer contour of binarized silhouette of a motion object and a bounding box placed around the silhouette are chosen as the basic image features called the distance vectors. The Spatio-temporal distance vectors are extracted using four view directions to the outer of the silhouette from the bounding box, they are top-, bottom-, left-, and right-views. Depending on the distance vectors, normalized correlation functions are produced for each view directions to analysis and to estimate the human motions, and to classify the motion objects. Experimental results on the different test image sequences demonstrate that the proposed algorithm has an encouraging performance with relatively robust and low computational cost.

1 Introduction

With increasing demands of visual surveillance systems, human action analysis has recently gained more interest in the research studies. The growth in this area is being driven by the increased availability of inexpensive computing power and image sensors [1] [2], and automatic human behavior understanding is also the other important applications [4]. There has been considerable interest in the area of human motion classification [18], tracking [3][13], analysis [15][10], and human identification based on gait analysis [7][2]. More references can be found in [8].

Real-world implementation for analyzing human motion and gait recognition for video applications will have to be computationally inexpensive and be applicable to real scenes in which objects are small and data is noisy. For the human motion analysis, using the geometrical shape models has the advantage that they have much information than directly obtained features. But the difficulty and cost of calculation in extracting the models from the input frames are disadvantage of using shape models for real time video surveillance applications. Those difficulties prevent researches from concentrating on cognition part of motion analysis process [9] [12]. The periodic nature of human walking has also been widely used in human motion analysis and in gait recognition and related application [11][14]. Several solutions have been proposed for measuring the periodicity of human motion analysis. The study in [15] presented a 3-D based detection in curvature space. The cyclic motion from optical flow domain was illustrated by the study in [14].

This paper presents a set of view independent techniques integrated into a low-cost PC based real time visual surveillance system for automatic gait cycle estimation, human motion classification and analysis their activities in monochromatic video. Periodical motion signatures obtained from view-based distance vectors are also a robust clue for view independent gait recognition. The novel approach presented is basically to produce view directions based 1-D distance vectors represent the differences between the outer counter of the binarized silhouette and the bounding box placed around the silhouette. Four 1-D signals are produced for each view, they are top-, bottom-, left-, and right-views. Then correlation-based a similarity function is produced for each view to estimate gait cycle of moving silhouette, to classify the motion objects, and then also to analysis human motions. The goal of the classification algorithm is to classify human and no-human based on their 1-D distance vectors data. Human motion analysis algorithm is also developed for detecting periodical gait cycle and for distinguishing walking and running actions.

2 Spatio-Temporal Human Motion Representation

Spatio-temporal human motion representation is based on the view directions to the 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 view directions to the silhouette. They are top-, bottom-, left- and right-view directions. Four distance vectors are produced from the view directions to the silhouette. The distance vectors are the differences between the bounding box placed around the silhouette and the outer counter of the silhouette, 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 the silhouette image *bs(t)*, as shown in figure 1.a. Each value $F_T(x,t)$ is then a count of the number of pixels between the top side of the bounding box of the object (person) and the outer counters in that columns x of the 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 pixels between the bottom side 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 pixels) from the left side of the bounding box to the outer counter of the 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 the silhouette as shown in figure 1.c.

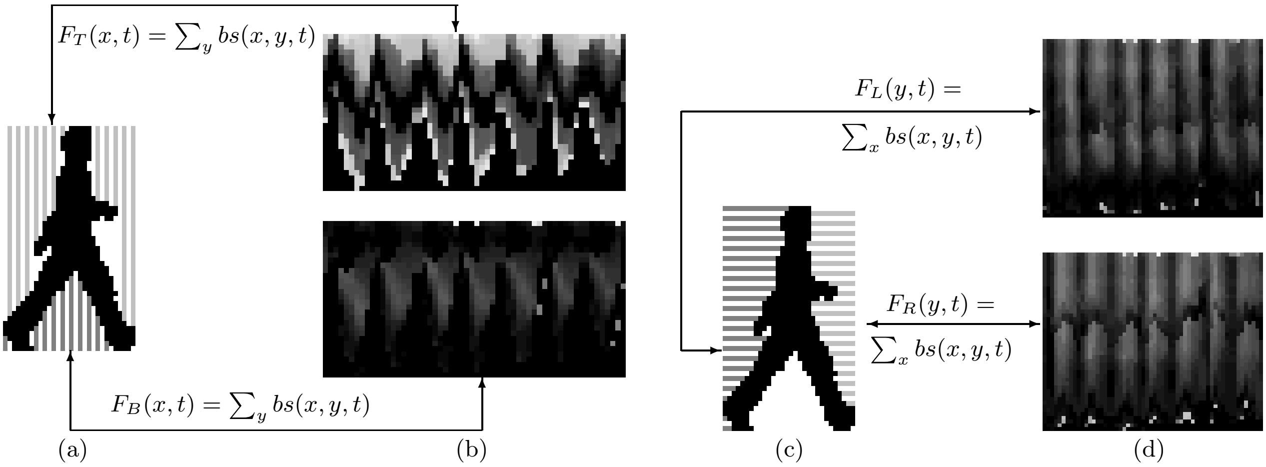


Fig. 1. Spatio-temporal motion representations. **(a)** Top and bottom view directions, **(b)** temporal plot of the distance vectors, **(c)** Left and right view directions, **(d)** temporal plot of the distance vectors.

The variation of each component of the each view direction distance vectors can be regarded as a silhouette signature of that object. 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 the silhouette. The brighter a pixel in figure 1.b and 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 a dynamic background frame estimated and updated in time, more details can be found in [5][6]. 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 the 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/tracking module gives a sequence of bounding boxes for every object. Reference signals $R_v(x/y)$ for v view direction distances at location column/row *(x/y)* to the silhouette are first obtained. The reference signals are produced from every previous frames of the current frames in the sequence. In other words, the $F_v(x/y, t-1)$ signals produced from previous frame are becoming the reference signals, $R_v(x/y)$, to the related view distance vectors in the current frame. Then the same manner is sequentially repeated for the following frames to update the reference signals. Then overall cost function is defined all definition view distance vectors of $F_v(x/y,t)$, where $F_v(x/y,t)$ is view direction distance vector for at location column/row (x/y) in time t for v

view direction, then the period is also defined as,

$$C_v(x/y) = \sum_{x=1}^{X}\sum_{y=1}^{Y} cor(R_v(x/y), F_v(x/y, n)) \tag{1}$$

$$period = argmax \sum_{x=1}^{X}\sum_{y=1}^{Y} cor(R_v(x/y), F_v(x/y, n)) \tag{2}$$

Where X is the width of the bounding box, Y is the height of the bounding box. The physical meaning of it is also to calculate the overall response of the signal $F_{v,t}$ with the reference signal $R_v = F_{v,t-1}$ at a given each view distance vector within the bounding box.

2.2 Periodicity Detection

Periodicity detection is especially very important at the gait identification. Because 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 [16][17][7]. Gait period analysis serves to determine the frequency and phase of each observed sequence so as to align sequences before matching. In [16], width time signal of the bounding box of moving silhouette derived from an image sequence is used to analyze 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 [17]. In [7], the aspect ratio of the bounding box of moving silhouette as a function of time is used to determine the period of the gait. Different from them, here we present four view distance 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.

Some experimental results on the normalized correlation process, as given in the equations 1-2, which is the first feature extraction step in the period analysis of each gait sequence are 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.b). In that, a reference signal for each view is constructed using the silhouette in previous frame in the image sequence. Thus, the reference signals are naturally normalized to eliminate the influence of spatial scale and signal lengths because of the basic idea presented here. The correlation results using four view distance vectors for the image sequence includes walking and running person's actions are grabbed on different views are acquired as shown in figure 2.b first and last two figures, respectively. From the correlation plot, we note 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 image sequence used in the experimental studies includes walking or running persons in lateral view, the normalized correlation results obtained

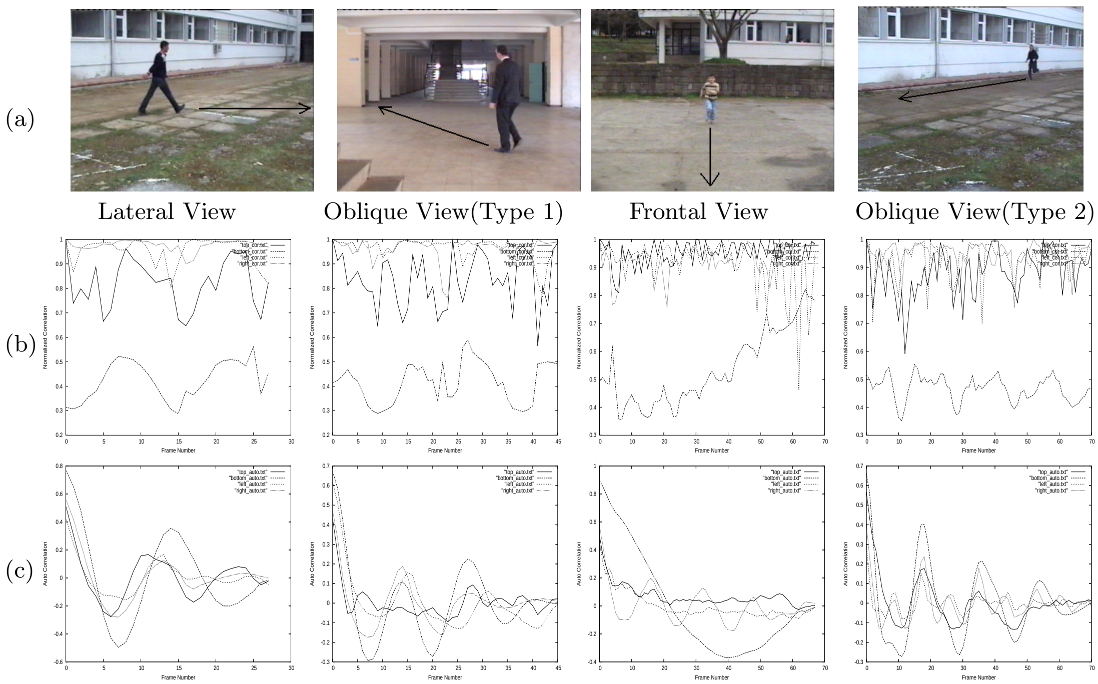
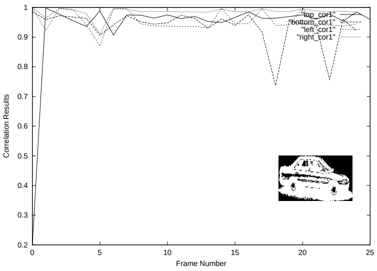


Fig. 2. Normalized Correlation results: (a) Some sample images in our database. (b) Normalized Correlation results for walking (first two figures) and for running person (last two figures). (c) Auto Correlation signals for cyclic detection.

from the all view distance vectors presented in this paper exhibit the periodical signals. For oblique views, the top- and bottom-views distance vectors based correlation has given more robust results than the left- and right-views distance vectors. The characteristics in their results on the test image sequences show the similar signal characteristics to the aspect ratio signals for lateral and oblique views, as shown in figure 2. For frontal views, the aspect ratio signals has not given a periodical signal on both actions (walking and running). The view distance vectors-based normalized correlation results have successfully exhibit the periodical signal characteristics. When the experimental results on the cyclic detection, as shown in figure 2.c, are especially considered, the robustness on the results is clearly seen on the processing of the autocorrelation. The more implementations on figure 2.c are given in section 4.

3 Motion Classification

In this section, our aim is to classify two type of objects: humans and vehicles by testing period existence in their motion. For the motion object classification, there are very nice parameters produced from silhouette structure and colors data in the bounding box placed around the motion object [1][13][18]. The novel



	Top View		Bottom View		Right View		Left View	
	Human	Vehicle	Human	Vehicle	Human	Vehicle	Human	Vehicle
Mean	0.504	0.965	0.577	0.935	0.772	0.982	0.753	0.956
A.C.D	0.274	0.021	0.262	0.052	0.173	0.016	0.104	0.030

Fig. 3. (Left) Correlation results on a car object in motion for each view direction distance vectors, **(Right)** Example data of the experiments on Mean and A.C.D values for the motion objects classification

idea in here is to present spatio temporal based parameters to classify the type of objects as human and vehicle. It is straightforward to think of mean, and averaging of cumulative differencing (ACD) on the results of correlation process performed on the distance vectors for decision since the combination of two characteristics is specific for human motion.

A neural network classifier is trained for each view-direction based classification data. The neural network is a standard three-layer network, trained using the back propagation algorithm. Input features to the network are measured directly from the mean and ACD values obtained by averaging of the results of correlations based on reference signals. Thresholds in each nodes in the neural net are the mean and the ACD. In our experiments, typical threshold values are 0.70 to 0.80 for the mean, 0.10 to 0.12 for the averaging of the cumulative differencing on the correlation results. There are three output classes in the network; human, vehicle, and clutter. When table in figure 3 is considered, it is clearly seen that the classification now becomes a simple process as thresholding spatio temporal parameters, which could be implemented with many fast and reliable methods. The mean and the ACD parameters and their averaging for two types of objects are typed in table in figure 3. Experimental results show that the parameters used can be good enough parameters to quite easily distinct those two types of objects. The thresholding values are determined by our experiments and correctly classify $\approx 90\%$ of the candidate object motions.

4 Motion Analysis

In this section, 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. Figure 2.b 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. Then, the algorithm computes their autocorrelation to find peaks, as shown in figure

2.c. Then, their first-order derivatives are computed to find peak positions by seeking the positive-to-negative zero-crossing points. Due to the bilateral symmetry of human gait, the autocorrelation will sometimes have minor peaks half way between each major peaks [16][7]. Hence, the real period as the average distance between each pair of consecutive major peaks are estimated.

As compared to the some results in figure 2.c, in the experiments for the gait sequence for frontal view, while the periodical characteristics of moving silhouettes are correctly detected by left- and right-view distance vectors based gait cycle estimation (as plotted in figure 2.c). The view distance based algorithm for gait cycle analysis presented here has achievement more robust experimental results than aspect ratio based gait cycle detection for frontal views. For lateral and oblique views, both the view distance (all four view distances vectors) and aspect ratio based algorithm can easily detect the gait cycles.

In addition, it is also useful to move into the Fourier domain to distinguish walking and running actions. The autocorrelation signals are transformed to the frequency domain. In the frequency, the major cyclic components of the cyclic point can be finally extracted from the power spectrum of this processed signal. According to our experimental results, the average walking frequency is 4.8 [Hz] and for running it is 6.225[Hz] for frame rate 15 fps. Decision for motion analysis is made using a simple feed forward neural network. In the network, the decision is simply obtained by thresholding in frequency. That is, the threshold values in frequency were used to discriminate the walking and running actions. At this work, typical threshold value was taken **5.5[Hz]** while frame rate was 15 fps. As long as the silhouette is not in big nosy, encouraging performance has been achieved. Then their values are implemented by the network to have more robust decision for motion analysis. In our experiments, the algorithm explained correctly classify 85% of the human motion analysis even the silhouette data is produced from in noisy environments, and unless the variations on human actions are more than normal variations limits on the actions, such as speedy walking.

5 Experimental Results

A video surveillance database is established for our experimental results. 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. At the database established for gait cycle detection and analysis, all subjects walk and run along a straight-line path at free cadences in four different views with respect to the image plane, as shown some images in figure 2. The database includes 17 subjects and two sequences for each action (walking, running) and for each viewing angle per subject. For motion classification, 3 image sequence obtained in our university campus, and 4 video sequence taken from the Computer Vision Lab. database in the University of Central Florida have been used as test sequences.

For the test sequences, the detection and tracking algorithm that we use is provided by the work in [5][6]. The outputs of its method are the objects within bounding boxes. This gives us great convenience for the view-direction distance vectors based experiments. Table in figure 3 shows the some experimental results of the basic parameters for motion object classification. In the classification, the algorithm correctly classified $\approx 90\%$ of the candidate object motions. The main problem with vehicle classification is that when vehicles are temporarily occluded for long times, and vehicles turning from moving direction, they are sometimes misclassified. Humans are also often misclassified as temporarily stable objects.

In human motion analysis, walking and running actions in the surveillance scene were only considered to differentiate from each other and to estimate the cycle of gait which is very important in gait based human identification. The test results 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. The shadow is important problem for the silhouette based motion identification and analysis because the structure shape of the silhouette may be fluctuated by the shadow types. However, the basic idea presented here is not depending on the shape of the silhouette, it is based on the spatio temporal variations on the silhouette in time. Intuitively, our algorithm will produce reliable results on motion objects even in strong shadow situations. This is one of the future work for testing the algorithm presented.

6 Conclusions

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, and the robustness to some factors such as motion object size and regular noise effects. In the motion object classification, a spatio temporal based parameters used is also the different novel idea then the others studies in literatures. Motion analysis and gait recognition implementations for real-world environment will have to be computationally inexpensive and be applicable to real scenes in which motion objects are small and data is noisy. The notion of using view direction vectors to classify and to analyze object motion is a useful one under these conditions. The results of the view independent cyclic detection have encouraged us to strongly say that the view distance vectors based spatio temporal data of silhouettes can involve the some characteristics to discriminate individuals by the way they walk.

7 Acknowledgment

This work is supported in part by Research Foundation of Karadeniz Technical University. (Grant No. KTU-2002.112.009.1).

References

1. R.T. Collins, A. J. Lipton, H. Fujiyoshi, T. Kanade.: Algorithms for Cooperative Multi sensor Surveillance. Proc. of IEEE Vol. 89. No.10, (2001).
2. G. V. Veres, L. Gordon, J.N. Carter, M. S. Nixon, *What image information is important in silhouette-based gait recognition?* Proc. of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'04, 2004.
3. P. Perez, C. Hue, J. Vermaak, M. Gangnet, *Color-Based Probabilistic Tracking,* Proc. of European Con. on Computer Vision, Copenhagen, 27 May- 2 June 2002, Denmark.
4. Mubarak Shah, *Understanding human behavior from motion imagery,* Mach. Vis. & Appl., Spec. Iss.: Human modeling, analysis, and synthesis, Vol.14, Iss.4, Sept.2003.
5. M. Ekinci, E. Gedikli, *A New Algorithm of Background Model Initialization and Maintenance for Real-Time Video Surveillance Applications.* To be Appear in Turkish Journal Electrical and Computer Science, June, 2005.
6. M. Ekinci, E. Gedikli, *Background Estimation Based People Detection and Tracking for Video Surveillance.* Springer LNCS 2869, ISCIS 2003, Computer and Information Sciences, 18*th* Int. Symp., pp. 421-429, Turkey, November, 2003.
7. L.Wang, T.Tan, H.Ning, W.Hu, *Silhouette Analysis-Based Gait Recognition for Human Identification* ,IEEE Trans. on Pat. Anal.&Mach. Intell., Vol.25, Dec., 2003.
8. L. Wang, W. Hu, T. Tan, *Recent developments in human motion analysis*, Pattern Recognition, Vol. 36, pp. 585-601, 2003.
9. R. Cutler, L. S. Davis, *Robust real-time periodic motion detection, analysis and applications* IEEE Trans. Pattern Anal. Machine Intell., Vol. 22, August 2000.
10. J.K. Aggarwal, Q. Cai, *Human Motion Analysis: A Review,* Computer Vision and Image Understanding, vol. 73, n. 3 pp. 428-440, March 1999.
11. Li B., Holstein H., *Recognition of Human periodic motion -a frequency domain approach* Proc. 16th Inter. Conf. on. Pattern Recognition, Vol.1, pp.311-314, 2002.
12. Yanxi Liu, R. T. Collins, T. Tsin, *Gait Sequence Analysis using Frieze Patterns*, Proc. of European Conf. on Computer Vision, 2002.
13. A. J. Lipton, H. Fujiyoshi. R. S. Patil, *Moving target classification and tracking from real-time video.* Proc., Workshop Appl. of Computer Vision, Princeton, Oct., 1998.
14. A. A. Efros, A. C. Berg, G. Mori, J. Malik, *Recognizing Action at a Distance*, Proc. Ninth IEEE Int. Conf. on Computer Vision, pp.726-733, 2003.
15. S. M. Seitz, C. R. Dyer, *View-Invariant Analysis of Cyclic Motion*, Int. Journal of Computer Vision, Vol. 25, pp.1-23, 1997.
16. C.BenAbdelkader, R.Cutler, L.Davis, *Stride and Cadence as Biometric in Automatic Person Identification and Verification* ,Int.Conf. Aut. Face & Gesture Rec., 2002.
17. 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.
18. O. Javed, S. Ali, M. Shah, *Online Detection and Classification of Moving Objects Using Progressively Improving Detectors*, IEEE, Proc. of Int. Conf. on Computer Vision and Pattern Recognition, CVPR'05, 2005.