MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
8-2

VIEW DISTANCE BASED HUMAN MOTION ANALYSIS

M. Ekinci, E. Gedikli, V. V. Nabiyev

Karadeniz Technical University, Department of Computer Engineering
61080, Trabzon Turkey

ABSTRACT

This paper presents a novel 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 distances 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. Correlation-based a similarity function in the time domain is calculated to classify the motion objects and a similarity function in the frequency domain is then also extracted to analysis human motions. 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

Human behavior understanding is to analyze and recognize the motion region segments by reason of human actions in frames and to produce high-level description of human actions [1, 3]. There has been considerable interest in the area of human motion classification-tracking [13], and analysis [2] in recent years. Further works are also focused to human identification based on gait analysis [6, 9]. Biometrics is also a technology that makes use of the physiological or behavioral characteristics to authenticate the identifies of people [4]. More references can be found in [7, 8]. Analyzing human motion for video applications is a complex problem. Real-world implementations 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 [10, 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 also illustrated in [14].

This paper presents a set of techniques integrated into a low-cost PC based real time visual surveillance system for simultaneously human motion classification and analysis their activities in monochromatic video. Periodical motion signatures obtained from view-based distance vectors are a robust clue for moving object classification. A new, flexible, simple but robust motion analysis algorithm using view directions-based 1-D distance vectors is presented in this study. The approach is basically to produce 1-D distance vectors represent the differences between the outer counter of the binarized silhouette and the bounding box placed around the silhouette. Thus, four 1-D signals are simultaneously produced for each view directions, they are top-, bottom-, left-, and right-views. Then correlation-based a similarity function in the time domain is produced for each view directions to classify the motion objects and a similarity function in the frequency domain is then also extracted 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 distinguishing walking and running actions. The key idea in this work is that simple, fast extraction of the broad internal motion features of an object can be employed to classify and to analyze its motion.

2. HUMAN MOTION REPRESENTATION

Spatio-temporal human motion representations based on view directions to the silhouette are generated from a sequence of a binary silhouette images *bs(t)= bs(x,y.t)*, indexed spatially by pixel location *(x,y)* and temporally by time *t*. The representations are based on four distances. The distances are encoded a measure of shape from the difference between the bounding box placed around the silhouette and the outer counter of the silhouette, as shown in figure 1. The distances are also chosen as the feature vectors. From a new 2D image $F_T(x,t) = \sum_y bs(x,y,t)$, where each column (indexed by time *t*) is the top-view distances of the silhouette image *bs(t)*, as shown in figure 1 top. Each value $F_T(x,t)$

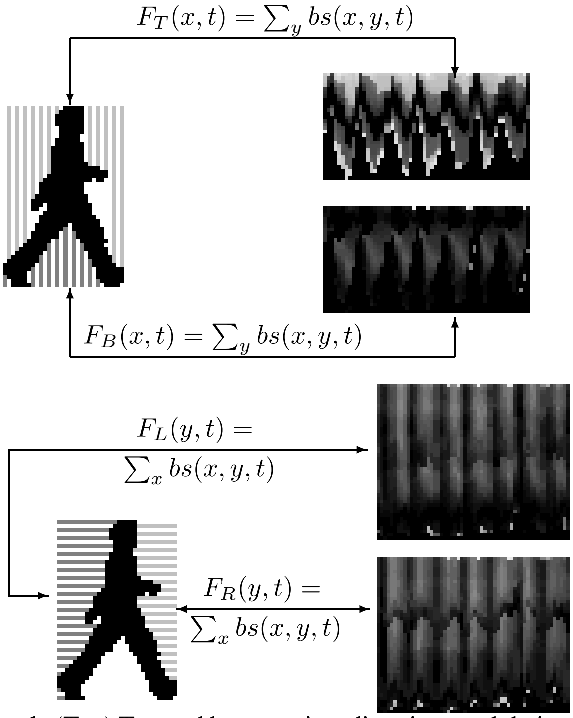

Fig. 1. (Top) Top and bottom view directions and their temporal plot of the distance vectors, (Bottom) Left and right view directions and their temporal plot of the distance vectors.

is then a distance along a given view direction (top-view direction) is computed as the distance vector in the locations of the top of the bounding box and the outer counters in that columns *x* of the silhouette image *bs(t)*. The results is a 2D pattern, formed by taking the distances from the top view direction together to form a spatio-temporal pattern. A second pattern produced from the bottom view direction $F_B(x,t) = \sum_y bs(x,y,t)$ can be constructed by taking the column distances from the bottom-view direction, as shown in figure 1 top, similarly and symmetrically to the $F_T(x,t)$.

The third pattern $F_L(y,t) = \sum_x bs(x,y,t)$ is then constructed by taking the row distances from the left-view direction to the outer counter of the silhouette. The last pattern $F_R(y,t) = \sum_x bs(x,y,t)$ is finally constructed by taking the row distances from the right-view direction, similarly and symmetrically with the third pattern, $F_L(y,t)$, as shown in figure 1 bottom.

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, 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]. 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. It is also the following assumptions are made; (1) the orientation and apparent size of the segmented objects do not change significantly during several periods (or do so periodically); (2) the frame rate is sufficiently fast for capturing the periodic motion.

3. FEATURES DERIVED FROM VIEW DISTANCES

The output of the detecting/tracking module gives a sequence of bounding boxes for every object. A reference signal $R_v(x,y)$ for each *v* view direction distance vectors at location *(x,y)* to the silhouette is produced from a frame arbitrary selected in the image sequence. There are four view directions, that is four reference signals, top-view $R_T(x,y)$, bottom-view $R_B(x,y)$, left-view $R_L(x,y)$, right-view $R_R(x,y)$. $R_T(x,y)$ is a distance vector at location (x,y) for top view direction in the reference frame considered, and so on. Then overall cost function is defined all definition distance vectors of $W_v(t,x,y)$, where $W_v(t,x,y)$ is distance vector for at location (x,y) in time *t* for *v* view direction.

$$C_v(x,y) = \sum_{x=1}^{X} \sum_{y=1}^{Y} cor(R_v(x,y), W_v(n,x,y)) \quad (1)$$

The physical meaning of it is to calculate the overall response of the signal W_v with the reference signal R_v at a location (x,y) and for a given direction *v* within the bounding box.

To detect the period more efficiently, a reference signal for each view is simultaneously constructed using the silhouette arbitrary chosen as a reference silhouette in the image sequence. Then the reference signal gained is correlated with the corresponding view distance vectors produced from the following frames in the image sequence. An example of the correlation results is shown in figure 2. The correlation results shown in figure 2 are obtained for a walking person, for a running person and for a car, respectively. The reference binary silhouette images arbitrary chosen are also shown in figure 2. From the correlation plot, we note that view feature vectors change with time as the person transits through a period of action, there is a high degree of correlation among the feature vectors across frames.

Reliable experimental results have been achieved when the algorithm depending on correlation process using single reference signal has been tested for motion classification and analysis. Nevertheless, instead of using single reference signal, we prefer three reference signals for each view direction so as to minimize the possible noisy effects on silhouette which is the reference signal constructed.

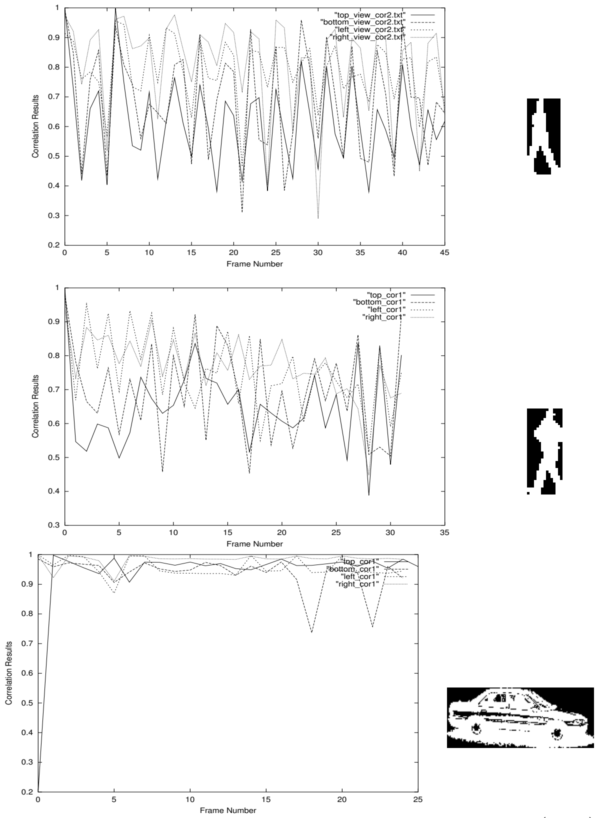

Fig. 2. Correlation results using a reference signal $R_v(x, y)$ with $W_v(t, x, y)$ for each view distance vectors, (Top) For a walking person, (Middle) For a running person. (Bottom) For a car. The right binary silhouette images are arbitrary chosen as reference silhouettes.

4. MOTION CLASSIFICATION

In this section, our aim is to classify two type of objects: humans and vehicles by testing period existence in their motion. The novel 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 three reference signals. In our experiments, typical values used as thresholds for the mean and the ACD to have a confident decision are 0.70 to 0.80 for the mean, 0.10 to 0.12 for the averaging of the cumulative differencing (ACD) on the correlation results. In order to have more robust motion classification, these parameters (Mean, ACD) are separately calculated for each correlation results achieved using three reference signals chosen. The averages of the mean and the ACD parameters calculated from three references-based correlation are simultaneously applied to the sub networks designed for each features (Mean and ACD). There are three output classes in the sub and full networks; human, vehicle, and clutter. In the sub-networks, a thresholds based classification is achieved. Then each results produced by the sub networks are applied to the last-layer of the network for final decision. 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 $\approx 95\%$ of the candidate object motions. It is also considered figure 2, the classification process can be easily clarified.

5. MOTION ANALYSIS

A low-level approach is presented to distinguish walking and running actions. Figure 2 displays a clear cyclical nature in the correlation results obtained on the distance vectors. To quantify these signals, it is useful to move into the Fourier domain. However, there is a great deal of signal noise, so a naive Fourier transform will not yet useful result. In the experiments, the power spectrum of the signal shows a great deal background noise. To emphasis the major cyclic component, an autocorrelation is performed on that results providing a new signal. But autocorrelation process introduces a new source of noise due to the bias of the correlated signal. Therefore, when low frequency components are autocorrelated, they remain in the signal and show up in the power spectrum as a large peak in the low frequencies. To reduce this problem, a high frequency pre-emphasis filter is applied to the signal before autocorrelation.

Humans in each image sequences perform walking or running movements toward different directions and the test motion sequences were also obtained from lateral view. Three reference signals in first three frames arbitrary chosen were simultaneously used as reference signals in the cycle detection processing to calculate the periods of human motions. 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. At this work, typical threshold value in the network was taken **5.5[Hz]** while frame rate was 15 fps. As long as the silhouette used for reference signals produced is not in strong noisy, encouraging performance has been achieved. Nonetheless, to minimize the possible distortion effects of the silhouette detected in noisy, three silhouettes in the following three frames in the image sequences are used to have three reference signals for each distance vectors. Therefore three sub-decision values are obtained for real-time human motion analysis on the three reference signals. Then their values are implemented by the next neural 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.

6. RESULTS AND CONCLUSIONS

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 and a CCD camera fixed on a pan-tilt motor platform are used to capture the video sequences. The algorithm presented here has been tried on a database of video sequences representative of situations which might be commonly encountered such as people walking, running, motion vehicles in the scene and frame rate was 15 frame per second. There are approximately 15 different video sequences include a mixture of adults, children and different vehicles as the motion objects.

For the test sequences, the detection and tracking algorithm that we use is provided by the work in [5]. The outputs of its method are the objects within bounding boxes. This gives us great convenience for the view-direction distance vectors based experiments. In the experimental studies for the classification algorithm applied to the test sequences, the main problem with vehicle classification is that when vehicles are temporarily occluded for long times, 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. 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.

The novelty in the presented algorithm lies in its simplicity, the efficiency of the implementation, the usefulness in real-time 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. The main disadvantage in this algorithm is to be directly influenced from any distortions long time on silhouettes in the sequence. However, a robust background estimation and upgrading algorithms will probably minimize to this problem [5].

7. ACKNOWLEDGMENT

This research work is supported by Karadeniz Technical University Research Foundation grant to (KTU-2002.112.009.1).

8. REFERENCES

[1] R.T. Collins, et.al: Algorithms for Cooperative Multi sensor Surveillance. Pr. of IEEE Vol.89 No.10, (2001).

[2] 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.

[3] Mubarak Shah, *Understanding human behavior from motion imagery,* Machine Vision and Applications, Special Issue: Human modeling, analysis, and synthesis, Vol. 14, Issue 4, Sept. 2003.

[4] A. K. Jain, A. Ross, S. Prabhakar, *An Introduction to Biometric Recognition*, IEEE Trans. on Circuit and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No.1, Jan. 2004.

[5] M. Ekinci, E. Gedikli, *Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance*. Turkish Journal Electric and Computer Engineering, Vol 10, No.1, January, 2005.

[6] 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.

[7] J.K. Aggarwal, Q. Cai, *Human Motion Analysis: A Review,* Computer Vision and Image Understanding, vol. 73, n. 3 pp. 428-440, March 1999.

[8] L. Wang, W. Hu, T. Tan, *Recent developments in human motion analysis*, Pattern Recognition, Vol. 36, pp. 585-601, 2003.

[9] B. Bhanu, J, Han, *Individual Recognition by Kinematics-Based Gait Analysis,* in Proc. of Int. Conf. on Pattern Recognition, pp. 343-346, 2002.

[10] R. Cutler, L. S. Davis, *Robust real-time periodic motion detection, analysis and applications* IEEE Trans. Pattern Analysis Machine Intell., Vol. 22, August 2000.

[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, October, 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.