Human Gait Recognition Using Temporal Slices

Shruti Srivastava and Shamik Sural

School of Information Technology
Indian Institute of Technology, Kharagpur, India
shrutisr@gmail.com, shamik@sit.iitkgp.ernet.in

Abstract. Gait along with body structure has been recognized as a potential biometric feature for identifying human beings. The spatial and temporal shape of motion of an individual is usually the same for all gait cycles and is considered to be unique to that individual. In this paper we introduce a Temporal Slice based approach for gait recognition. Temporal Slices are a set of two-dimensional images extracted along the time dimension of an image volume. They encode a rich set of visual patterns for similarity measure and have been widely used for motion detection. We show that the features obtained from tensor histogram of these temporal slices can be efficiently used as gait features for recognition of human beings.

Keywords: Gait biometrics, Temporal Slices, Tensor Histogram, Multiclass SVM.

1 Introduction

Recognition and verification of human beings is a major security concern in many restricted areas.There are various methods used to address this problem including use of signatures and passwords, biometric feature based methods like facial recognition, iris scan, fingerprints, etc. But none of these methods can be hidden from the suspects or people being monitored. Rather, they require cooperation from the subject in some way or the other. Till now the only perceivable biometric feature that can be captured from a distance is the gait. People walking in a passage can be easily monitored by hidden cameras, that too without hindering the normal activities in that arena.

Gait is the style of walking of a person. It can be termed as the shape of motion. Gait, together with the body structure, can be used as an efficient biometric feature for uniquely identifying a person. In fact, early medical studies show that there are 24 different components to human gait, and that, if all the measurements are considered, the gait of an individual is unique. Statistics show that if there are no external factors involved, the possibility of two individuals having the same pattern formed by their gait and body structure is also less [1,2]. External factors that can affect gait includes prior knowledge of being monitored, voluntary attempt to copy other's style of walking, injury, pregnancy, etc.

The gait of a person is a periodic activity with each gait cycle covering two strides; the left foot forward and the right foot forward. Gait biometrics can be

A. Ghosh, R.K. De, and S.K. Pal (Eds.): PReMI 2007, LNCS 4815, pp. 592–599, 2007.

considered to be derived from the shape and dynamics of the two strides. The gait dynamics is vulnerable to change under varying factors like walking surface, walking speed, etc. So, gait dynamics alone cannot be considered as a stable source of biometric information [3]. Both the shape and the dynamics of a gait cycle for a person forms the gait biometric feature for that person.

Gait volume is a collection of image sequences of an individual in a gait cycle. Temporal slices provide rich visual pattern and have so far been used for motion characterization. It can also have the potential of characterizing human motion as well as distinguishing one subject from another.

2 Related Work

Current gait recognition approaches may be classified into two main classes [1,2,4,5] Model based Approach and Motion based Approach. Both methodologies follow the general framework of feature extraction, feature correspondence and high-level processing. The major difference is with regard to feature correspondence between two consecutive frames. Model based approach aims at modeling human body structure or motion and extracts image features from these models. They generally perform model matching in each frame of a walking sequence to match the model parameters such as trajectories, limb length and angular speed [5]. Motion based or appearance based approach characterizes the whole body movement pattern by a compact representation regardless of the underlying structure. It can be further classified into state-space method and spatio-temporal method [2]. State space method considers gait motion to be composed of a sequence of static body poses and recognizes it by considering temporal variations of observations with respect to that static pose. Spatio-temporal method characterizes the spatio temporal distribution generated by gait motion in its continuum.

Liu and Sarkar [3] categorized gait recognition approaches into the following three types: i) Temporal Alignment Based ii) Static Parameter Based and iii) Silhouette Shape Based. The temporal alignment based approach emphasizes both shape and dynamics. It aims at extracting features from silhouettes, pre-shape representation, silhouette parts and Fourier descriptors. A sequence of these features is aligned with the sequence to be matched by either simple temporal correlation [2], dynamic time warping [6], hidden Markov models [4], phase locked-loops or Fourier analysis. The distance measure is either Euclidean [2], simple dot product-based or Procrustes distance. The Static Parameter Based approach uses parameters to characterize gait dynamics, such as stride length, cadence, stride speed and sometimes static body parameters like ratio of various body parts. The Silhouette Shape Based approach emphasizes on silhouette shape similarity and disregards temporal information. This approach generally transforms the silhouette sequence into a single image representation such as averaged silhouette [5] or an image representation derived from the width vectors in each frame (Freize patterns).

Sagawa et al. [8] applied fourier transforms on the sliced planes of image volume to extract the frequency characteristics of gait for an individual. They used these frequency characteristics for tracking the moving individuals in a video shot. Ngo [7] used temporal histogram to find the distribution of local orientation in the temporal slices obtained from the image volume of a video shot. The feature vector obtained from orientation histogram was used for clustering and retrieval of video shots.

3 Temporal Slice Based Approach

Our approach is based on the analysis and processing of gait patterns in temporal slice images. A temporal slice is a set of two-dimensional (2-D) images extracted along the time dimension of an image volume.The gait volume is the set of temporal images in a gait cycle and the temporal slices are extracted from this gait volume. One dimension of the temporal slice is time and another dimension is the x or y axis of the spatial frames in a gait volume.

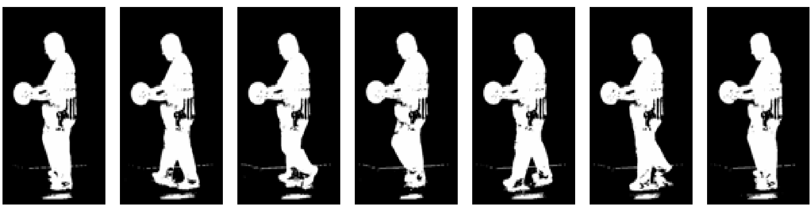

Fig. 1. Sequence of Frames in a Gait cycle (here, N=7)

If there are N number of frames in a gait cycle and the image size is height $\times$ width (Figure 1), the gait volume will be a 3D representation of dimension height $\times$ width $\times$ N (Figure 2). Thus, gait volume has one dimension as time while the other dimension has 2D frame sequences in the spatial domain. Figure 2 represents a gait volume of seven frame sequences in time domain. The horizontal slice (Figure 3 (a)) has one dimension as the time and another dimension as x axis (i.e. y value is fixed). In vertical slice (Figure 3 (b)) dimensions are time and y axis (height), value of x (width) is fixed.

We apply Gaussian derivatives of the horizontal slice to obtain the partial derivatives along the x axis ($\mathbf{H_x}$) and the partial derivatives along the time axis ($\mathbf{H_t}$). Using these gradient vectors we calculate the gradient tensor as follows:

$$Q = \begin{pmatrix} \mathbf{H_x} \\ \mathbf{H_t} \end{pmatrix} \begin{pmatrix} \mathbf{H_x} & \mathbf{H_t} \end{pmatrix} \tag{1}$$

$$= \begin{pmatrix} \mathbf{H_x}^2 & \mathbf{H_x H_t} \\ \mathbf{H_x H_t} & \mathbf{H_x}^2 \end{pmatrix}$$

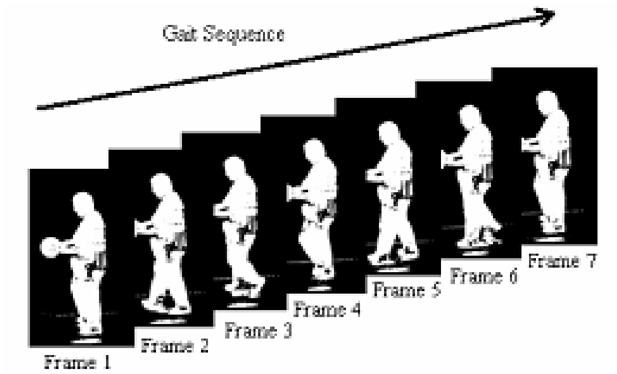

Fig. 2. Gait Volume

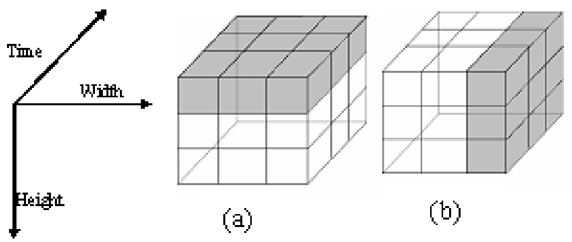

Fig. 3. (a) Horizontal Slice (b) Vertical Slice of a gait volume

Which can be written as:

$$Q = \begin{pmatrix} \mathbf{q_{11}} & \mathbf{q_{12}} \\ \mathbf{q_{21}} & \mathbf{q_{22}} \end{pmatrix} \tag{2}$$

Here, H_x and H_t are multiplied pixel by pixel, i.e. ${H_x}^2(i,j) = H_x(i,j) * H_x(i,j)$ and so on.

After calculating gradient tensor for all the horizontal slices, we get a gradient tensor cube. The gradient tensor cube consists of three cubes $\mathbf{q_{11}}, \mathbf{q_{12}}, \mathbf{q_{22}}$ each of the same dimension as that of the gait volume.

On this gradient tensor cube, we use Sobel averaging over a 3x3 window (on each of $\mathbf{q_{11}}, \mathbf{q_{12}}, \mathbf{q_{22}}$ separately to get $\mathbf{S_{11}}, \mathbf{S_{12}}, \mathbf{S_{22}}$). The averaging is done in spatial domain, i.e. by fixing the time value. This gives us the structure tensor (τ).

$$\tau = \begin{pmatrix} \mathbf{S_{11}} & \mathbf{S_{12}} \\ \mathbf{S_{21}} & \mathbf{S_{22}} \end{pmatrix} \tag{3}$$

Here, $\mathbf{S_{11}}, \mathbf{S_{12}}, \mathbf{S_{22}}$ are the 3D cubes of the same size as that of the gait volume.

The Sobel averaging is done for edge detection; $\mathbf{S_{11}}$ gives orientation in x direction, $\mathbf{S_{22}}$ in time direction and $\mathbf{S_{12}}$ in both the directions.

The angle of rotation of the principle axes of the structure tensor for each pixel is given by:

$$\theta = \frac{1}{2} \arctan \frac{2\mathbf{S_{12}}}{\mathbf{S_{11}} - \mathbf{S_{22}}} \tag{4}$$

The local orientation of a pixel can be written as:

$$\phi = \begin{cases} \theta - \frac{\pi}{2} & \theta > 0 \\ \theta + \frac{\pi}{2} & \text{otherwise} \end{cases}$$

Thus, $\phi = \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$

The certainty factor with which a pixel has a particular local orientation is given by:

$$c = \frac{(\mathbf{S_{11}} - \mathbf{S_{22}})^2 + 4\mathbf{S_{12}}^2}{(\mathbf{S_{11}} + \mathbf{S_{22}})^2} \tag{5}$$

After getting the orientation angle and the certainty factor for each pixel, the orientation angles are quantized. Here we quantize the orientation angle into 8 levels between $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$.

All the pixels having the same orientation are grouped together and their certainty factors are added. This is done for each frame to obtain a 1D orientation histogram for each frame. This 1D orientation is summed across the time domain to give a vector (equal to the number of quantization levels of orientation angle), which constitutes a feature vector for the horizontal slices.

Similar procedure is applied on the vertical slices to get another set of features. The feature vector from horizontal slices and the feature vector from vertical slices are clubbed together to give the feature vector of the gait of a given individual.

We use Multiclass Support Vector Machine (SVM)[15] for training and testing our model. SVM uses discriminating method of training and predicts multivariate outputs. Let S = $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m)\}$ be a set of m training examples. $\boldsymbol{x}_i$ is a vector of the training dataset ($\boldsymbol{x}_i \in \chi$) and y_i is an integer from the set $Y = \{1, \ldots, k\}$, where k is the number of classes. A multiclass classifier is a function H: $\chi \rightarrow Y$ that maps an instance $\boldsymbol{x}$ to an element y of Y. The classifier used by us is of the form:

$$H_M(\boldsymbol{x}) = arg \max_{r=1}^{k} \{\boldsymbol{M}_r . \boldsymbol{x}\} \tag{6}$$

The matrix M is of size k×n, n is the number of features in the vector, $\boldsymbol{M}_r$ is the rth row of M. The predicted label for a probe vector $\boldsymbol{x}$ is the index of the row attaining the highest similarity score with $\boldsymbol{x}$. The empirical error is given by:

$$\epsilon_S(M) = \frac{1}{m} \sum_{i=1}^{m} [\max_r \{\boldsymbol{M}_r . \boldsymbol{x} + 1 - \delta_{y,r}\} - \boldsymbol{M}_y . \boldsymbol{x}] \tag{7}$$

$\delta_{p.q} = 1$ if p=q and 0 otherwise. This bound is zero if the confidence value of the correct label is larger by at least 1 than the confidence assigned to rest of the labels. Otherwise there is a loss which is linearly proportional to the difference between the confidence of the correct label and the maximum among the confidence of other labels.

4 Result and Conclusion

We have used CMU's Mobo dataset [9] for our experiments. It comprises of indoor video sequences of 25 subjects. Different modes of walking are considered i.e. walking on an inclined plane, walking with a ball in one hand, fast walk and slow walk. A person in different modes of walk is considered to be of different class. So, if there are five people in our dataset each recorded for all four modes of walk, there is a total of twenty classes. The binary silhouettes in the dataset are normalized and horizontally aligned. By normalization we mean to proportionally resize each silhouette image so that all silhouettes have the same height. Horizontal alignment is to center the upper half silhouette part with respect to its horizontal centroid.

The CMU dataset contains binary silhouettes of size 640×486 and there are 34 such frames per gait cycle. So, the dimension of gait volume is 640×486×34. For horizontal slices we fix the first dimension, while for the vertical slices we fix the second dimension. Thus, the dimension of a horizontal slice is 486×34 while that for vertical slice is 640×34.

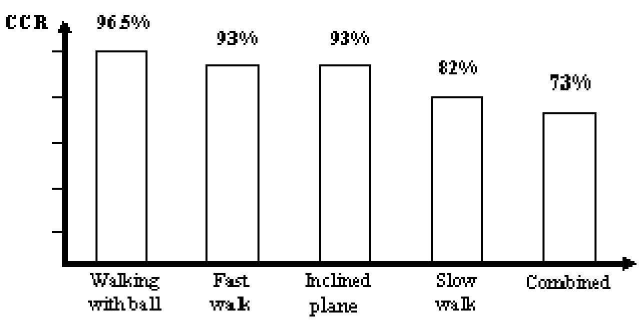

Fig. 4. The Correct Class Recognition (CCR) percentage

We have used 10 fold cross validation technique for testing our model. We tested on all the four types of walk separately. The correct class recognition (CCR) for individuals walking with a ball was 96.5%. For individuals walking at high pace and for individuals walking on an inclined plane, the CCR was 93%. CCR for slow walk was 82%. When our training and testing data contained feature vectors from all the modes of walking, the CCR achieved was 73% (Figure 4).

Algorithm Used	Slow walk	With Ball	Fast walk	Inclined plane
CMU [11]	100	92	76	-
UMD [12]	72	-	32	-
UMD [13]	72	-	12	-
Georgia Tech*	-	50	45	-
MIT [14]	100	50	64	-
Baseline [10]	92	88	72	-
Temporal Slice	**82**	**96.5**	**93**	**93**

* http://www.cc.gatech.edu/cpl/projects/hid/CMUexpt.html

\- result unknown

Fig. 5. Comparative Results in CCR of different Algorithms on CMU Dataset (Ref. [10])

A comparative result is shown in Figure 5. It gives the CCR of various algorithms tested on CMU Mobo dataset. We have taken this data from Sarkar's work [10] and included our result. Our algorithm performs better than others for fast walk and walking with a ball. There was no data available for inclined plane walk, so a comparison could not be done. Thus, temporal slice can be potentially used as a feature for gait recognition.

References

1. Wang, L., Tan, T., Ning, H., Hu, W.: Automatic Gait Recognition Based on Statistical Shape Analysis. IEEE Transactions on Image Processing 12(9), 1120–1131 (2003)
2. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette Analysis Based Gait Recognition for Human Identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12), 1505–1518 (2003)
3. Liu, Z., Sarkar, S.: Improved Gait Recognition by Gait Dynamics Normalization. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(6), 863–876 (2006)
4. Kale, A., Sundaresan, A., Rajagopalan, A.N., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., Chellappa, R.: Identification of Humans Using Gait. IEEE Transactions on Image Processing 13, 1163–1173 (2004)
5. Han, J., Bhanu, B.: Individual Recognition using Gait Energy Image. IEEE Transactions of Pattern Recognition and Machine Intelligence 28(2), 316–322 (2006)
6. Bolgouris, N.V., Plataniotis, K.N., Hatzinakos, D.: Gait Recognition using Dynamic Time Warping. In: Sixth IEEE Workshop on Multimedia Signal Processing, pp. 969–979 (2004)
7. Ngo, C.W., Pong, T.C., Zhang, H.J.: On Clustering and Retrieval of Video Shots through Temporal Slice Analysis. IEEE Transactions on Multimedia 4(4), 446–458 (2002)
8. Sagawa, R., Makihara, Y., Echigo, T., Yagi, Y.: Matching Gait Image Sequences in the Frequency Domain for Tracking People at a Distance. In: Proceedings of 7th Asian Conference on Computer Vision, vol. 2, pp. 141–150 (2006)

9. Gross, R., Shi, J.: The CMU Motion of Body (MoBo) Database, Tech. Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University (2001)
10. Sarkar, S., Phillips, P.J., Liu, Z., Robledo-Vega, I., Grother, P., Bowyer, K.W.: The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2), 162–177 (2005)
11. Collins, R., Gross, R., Shi, J.: Silhouette-Based Human Identification from Body Shape and Gait. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 366–371 (2002)
12. Kale, A., Rajagopalan, A., Cuntoor, N., Krueger, V.: Gait-Based Recognition of Humans using Continuous HMMs. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 321–326 (May 2002)
13. Ben Abdelkader, C., Cutler, R., Davis, L.: Motion-Based Recognition of People in Eigengait Space. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 267–272 (2002)
14. Lee, L., Grimson, W.: Gait Analysis for Recognition and Classification. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 155–162 (2002)
15. Crammer, K., Singer, Y.: On the Algorithmic Implementation of Multiclass kernel based Vector Machine. Journal of Machine learning Research, 265–292 (2001)