Person identification from spatio-temporal volumes

Yumi Iwashita[1], Maria Petrou[2]

[1]Information Science and Electrical Engineering, Kyushu University, Japan.
[2]Electrical and Electronic Engineering, Imperial College London, UK.
Email: yumi@is.kyushu-u.ac.jp

Abstract

This paper proposes a novel identification technique for a person from gait and body shape. Although the shape of one's body has not been considered much as a characteristic for the person identification, it is closely related to gait and it is difficult to disassociate them. The proposed technique utilizes the full spatio-temporal volume carved by a person who walks and the average image created from the spatio-temporal volume. Affine moment invariants are derived from the spatio-temporal volume and the average image, and classified by a support vector machine. Experiments using a standard gait database show this method may produce better results than those based on gait analysis alone and k-nearest neighbor classification.

Keywords: Biometrics, person identification, affine moment invariants

1 Introduction

Biometric research encompasses a large number of characteristics of a person in order to identify them. These include physiological biometrics, related to the shape of the body, the oldest of which are the fingerprints; and the behavioral biometrics, related to the behavior of a person, the first of which used was the signature.

One of the recently introduced human characteristics that may be used for person identification is gait [1]. Gait is the peculiar way one walks and is a complex spatio-temporal biometric.

One of the other characteristics that has not been considered much is the shape of one's body. This is probably because the shape of the body changes with time and with clothing. However, new sensors are being developed now which can see through clothing. In addition, it is difficult to disassociate body shape from gait. So, this paper introduces a new hybrid biometric, combining body shape (physiological) and gait (behavioral). When a person walks, she/he carves a specific volume in the spatio-temporal domain. It is the shape of this volume we wish to consider as a biometric. In addition to this biometric, we consider another biometric, that is, the average image obtained from the spatio-temporal volume. In this paper we propose a novel person identification method which utilizes both the spatio-temporal volume and the average image.

The work that is closest to ours is that of Shah et al [2] who used the spatio-temporal volumes carved by a moving human limb to identify the action of a person.

This paper is organized as follows. Section 2 is a brief literature survey on the identification of individuals using gait. Research that considers the spatio-temporal shapes created by a walking person is reviewed in section 2.2. However none of these papers considers the full volumes as potential biometrics. Section 3 describes the methodology we shall use in this paper. Section 4 describes the data we shall use and the experiments performed. Our results and conclusions are presented in section 5.

2 Literature survey on gait as a biometric

Gait recognition has the advantage of being unobtrusive because body-invasive sensing is not needed to capture gait information. Moreover, gait recognition has the extra advantage that it may be performed from a distance. Several approaches have been proposed for the identification of a person from their gait. They may be mostly classified into two classes, model-based and appearance-based approaches.

978-1-4244-2582-2/08/$25.00 ©2008 IEEE

2.1 Model-based approaches

A model-based approach recovers explicit features describing gait dynamics, such as stride dimensions and the kinematics of joint angles. Bouchrika and Nixon [3] described spatial model templates for human gait in a parameterized form using Fourier descriptors. The positions of the joints of walking people were extracted by using the Hough Transform.

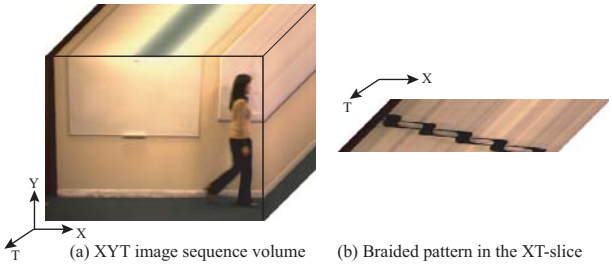
Cunado et al. [4] extracted the motion of the thigh, and defined their gait signature by Fourier analysis. Yam et al. [5] extended the system [4] to handle walking as well as running people. They extracted the motion of the hip, the thigh, and the lower leg by temporal template matching. Phase-weighted Fourier description gait signatures were derived from the extracted movements.

Urtasun and Fua [6] introduced a model composed of implicit surfaces attached to an articulated skeleton. The motion of people was tracked by using the 3D model and clusters of 3D points captured by a stereo camera. By using the 3D model, they increased the robustness to a changing view direction. Lee and Grimson [7] introduced gait representation based on moments extracted from orthogonal view video silhouettes. Their gait appearance feature vector comprised parameters of moment features of image regions. Seven ellipses were fitted to different parts of the binarized silhouette of the person and the parameters of these ellipses, such as location of their centroids, eccentricities, etc., were used as features to represent the gait of the person. BenAbdelkader et al. [8] used stride and cadence for the identification of people. The person's identity was defined based on parametric Bayesian classification of the cadence and stride feature vector. From their experiments, the variation in stride length with cadence was found to be linear and unique for different people.

2.2 Appearance-based approaches

Appearance-based approaches directly extract parameters from images without assuming a model of the human body and its motion. This approach characterizes body movement by the statistics of the spatio-temporal (XYT) patterns generated in the image sequences by the walking person. Here, the XYT patterns are formed by piling up frames in an image sequence as shown in Fig.1. There are many ways of extracting XYT patterns from the image sequences of a walking person.

The simplest approach is to use the sequence of binary silhouettes spanning one gait cycle and scaled to a certain uniform size [9]. Murase and Sakai [9] proposed a template matching method in the parametric eigenspace that was created from images.



(a) XYT image sequence volume (b) Braided pattern in the XT-slice taken at the ankle

Figure 1: XYT image sequence volume.

Sarker et al. [10] proposed a baseline algorithm of gait recognition, which estimated the silhouettes by background subtraction and performed recognition by spatio-temporal correlation of silhouettes. Collins et al. [11] used silhouettes corresponding to certain gait poses only. Liu et al. [12] adopted dynamics-normalized shape cues with a population HMM which emphasize difference in stance shapes between subjects and suppresses differences for the same subject under different conditions.

Other methods use a signature of the silhouette by collapsing the XYT data into a more terse 1D or 2D signal, such as vertical projection histograms (XT), and horizontal projection histograms (YT) [13]. Niyogi and Adelson [13] extracted XT sheets that encoded the person's inner and outer bounding contours detected by fitting 'snakes'. Similarly, Liu et al. [14] extracted the XT and YT projections of the binary silhouettes. They used a frieze pattern to represent gait motion, that is a pattern created by summing up the white pixels of a binarized image of a gait along the rows and columns of an image. BenAbdelkader et al. [15] characterized gait in terms of a 2D signature computed directly from the sequence of silhouettes. The signature consisted of self-similarity plots (SSP), defined as the correlation of all pairs of images in the sequence. Han et al. used the gait energy image (GEI) which is based on the average image [16].

Little and Boyd [17] used optical flow instead of binary silhouettes. They fitted an ellipse to the dense optical flow of the person's motion, then computed thirteen scalar features consisting of first- and second-order moments of this ellipse. Twelve measurements were provided from thirteen features.

3 Methodology

3.1 Two biometrics from human walking sequences

In this section, we describe the methodology we use in this paper. Here, we assume that a target region in an image sequence is extracted. Figure 1 (a) shows the 3D volume in the spatio-temporal (XYT) domain, and it is formed by piling up the

target region in the image sequences of one gait cycle, which is used to partition the sequences for the 3D volume. One gait cycle is a fundamental unit to describe the gait during ambulation, which occurs from the time when the heel of one foot strikes the ground to the time at which the same foot contacts the ground again. In this paper, we assume that the 3D volume consists of a number of small voxels.

The average image $I_{average}(x, y)$ is defined as follows:

$$I_{average}(x, y) = \frac{1}{N} \sum_{t=1}^{N} I(x, y, t), \qquad (1)$$

where N is the number of frames in one gait cycle and $I(x, y, t)$ represents the density of the voxels at time t. For characterizing these 2D average images and 3D volumes, we consider the 2D and 3D affine moment invariants as features.

3.2 2D and 3D affine moment invariants

In this section, we introduce 2D and 3D affine moment invariants. Affine moment invariants are moment-based descriptors, which are invariant under a general affine transform. The moments describe shape properties of an object as it appears. For an image $I(x, y)$ the 2D moment of order $(p+q)$ of an object O is given by

$$\mu_{pq} = \iint_{(x,y)\in O} x^p y^q I(x, y) dx dy. \qquad (2)$$

The discrete version of Eq.2 is written as

$$\mu_{pq} = \sum\sum_{(x,y)\in O} x^p y^q I(x, y). \qquad (3)$$

The center of gravity of an object in the image can be determined from the zeroth and the first-order moments by

$$x_g = \frac{\mu_{10}}{\mu_{00}}, \qquad y_g = \frac{\mu_{01}}{\mu_{00}}. \qquad (4)$$

Centralized moments are computed by using the coordinates x_g and y_g:

$$\mu_{pq} = \sum\sum_{(x,y)\in O} (x - x_g)^p (y - y_g)^q I(x, y). \quad (5)$$

Six affine moment invariants are listed below [18].

$$\begin{aligned}
I_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2) \\
I_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 \\
&\quad +4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
I_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} \\
&\quad -\mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \\
I_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} \\
&\quad -6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} + 9\mu_{20}^2\mu_{02}\mu_{12}^2 \\
&\quad +12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} \\
&\quad -18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 8\mu_{11}^3\mu_{30}\mu_{03} \\
&\quad -6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad +12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} \\
&\quad +\mu_{02}^3\mu_{30}^2) \\
I_5 &= \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) \\
I_6 &= \frac{1}{\mu_{00}^9}(\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{22}\mu_{13} \\
&\quad -\mu_{40}\mu_{13}^2 - \mu_{04}\mu_{13}^2 - \mu_{22}^3) \qquad (6)
\end{aligned}$$

For a 3D space the 3D moment of order $(p+q+r)$ of a 3D object O is given by the same procedure with 2D centralized moments.

$$\begin{aligned}
\mu_{pqr} &= \sum\sum\sum_{(x,y,t)\in O} \\
&\quad (x - x_g)^p (y - y_g)^q (t - t_g)^r I(x, y, t), \qquad (7)
\end{aligned}$$

where x_g, y_g and t_g are the coordinates of the center of gravity of an object in the 3D space.

Six 3D affine moment invariants are given in [19] [20], and two of them are listed below. For the rest of them, refer to [20] because of their long formulae.

$$\begin{aligned}
I_1 &= \frac{1}{\mu_{000}^5}(\mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} \\
&\quad -\mu_{200}\mu_{011}^2 - \mu_{020}\mu_{101}^2 - \mu_{002}\mu_{110}^2) \\
I_2 &= \frac{1}{\mu_{000}^7}(\mu_{400}(\mu_{040}\mu_{004} + 3\mu_{022}^2 - 4\mu_{013}\mu_{031}) \\
&\quad +3\mu_{202}(\mu_{040}\mu_{202} - 4\mu_{112}\mu_{130} + 4\mu_{121}^2) \\
&\quad +12\mu_{211}(\mu_{022}\mu_{211} + \mu_{103}\mu_{130} - \mu_{031}\mu_{202} \\
&\quad -\mu_{112}\mu_{121}) + 4\mu_{310}(\mu_{031}\mu_{103} - \mu_{004}\mu_{130} \\
&\quad +3\mu_{013}\mu_{121} - 3\mu_{022}\mu_{112}) + 3\mu_{220}(\mu_{004}\mu_{220} \\
&\quad +2\mu_{022}\mu_{202} + 4\mu_{112}^2 - 4\mu_{013}\mu_{211} - 4\mu_{121}\mu_{103}) \\
&\quad +4\mu_{301}(\mu_{013}\mu_{130} - \mu_{040}\mu_{103} + 3\mu_{031}\mu_{112} \\
&\quad -3\mu_{022}\mu_{121})) \qquad (8)
\end{aligned}$$

In the proposed method, we extract six 2D affine moment invariants from the average image and six 3D affine moment invariants from the spatio-temporal volume of each subject at first, and the classifier is trained by using the training data sets. Then in the identification phase, the same affine moment invariants are extracted from the average image and the spatio-temporal volume, and the subject is identified by the classifier.

4 Experiments

In this section we describe the experiments. In our experiments, we used a gait database collected by the University of Southampton [21]. The database contains raw image sequences and foreground masks. Figure 2 shows an example entry from the database. In our experiments, the support vector machine (SVM) was applied to the affine moment invariants as the classifier. The classification with the SVM, which is extended to the multiclass case, is performed using the Pattern Classification Program (PCP) [22], and the kernel type is the Radial Basis Function. We used the leave-one-out cross validation to estimate the classification error rate. The database contains 140 video sequences, which contain 20 different subjects with 7 sequences for every subject.

We carried out four experiments. In the first experiment the 3D affine moment invariants, which are calculated from the binary spatio-temporal volume, are used for the classification. Figure 3 shows the first two affine moment invariants used for 10 out of the 20 subjects and their whitened affine moment invariants. Here, data-specific variation can be removed by whitening. The resultant correct classification rate using whitened data was 75 %.

In the second experiment we used the 2D affine moment invariants of the average images for classification. Figure 4 (a) shows an example of the average images. Here, the images are properly aligned and scaled to a uniform height before calculating the average image. The resultant correct classification rate was 92 %. From this experiment, we could say that the average images show better performance for each person than the binary spatio-temporal volumes.

So in the next experiment, for emphasizing the difference of features in individuals, we created new images by subtracting the average image from the original images (Fig.4 (b)), and then the new spatio-temporal volume was formed by pilling up the target region in the new images of one gait cycle. Here, the average image was created from all training images. The resultant correct classification rate using the 3D affine moment invariants of the new spatio-temporal volume was 84 %, which was higher than the binary spatio-temporal volume. In the final experiment, we used 2D affine moment invariants of the average image and 3D affine moment invariants of the new spatio-temporal volume. The resultant correct classification rate was 94 %, which was the highest score in the series of the experiments.

Using the same database, Nixon et al. [3] used dynamic and static gait features to yield a feature vector. Static features include the body height, stride and heights of different body parts while dynamic features are the phase-weighted magnitudes of the Fourier frequencies for the hip and knee angular motions. The gait signature is derived using the adaptive forward floating search algorithm via selecting the features with higher discriminability values. They used the k-nearest neighbor rule as a classifier, and their system achieved a correct classification rate of 92 % using the leave-one-out cross validation rule. The comparison of their experiment and our experiment is shown in Table 1.

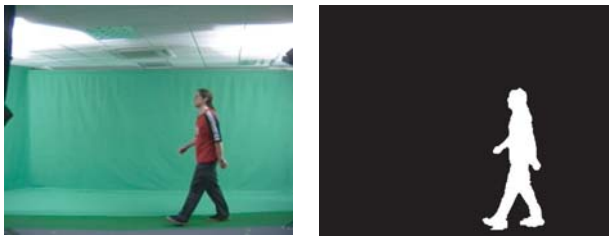
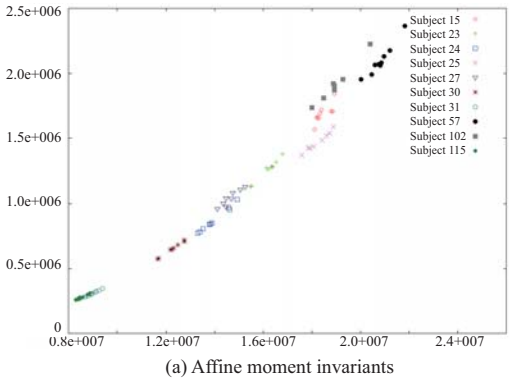
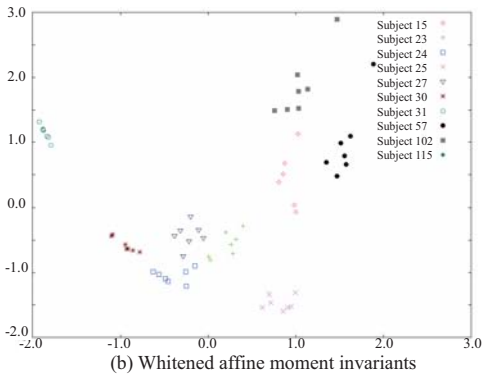


Figure 2: Samples from the University of Southampton database. [21]



(a) Affine moment invariants



(b) Whitened affine moment invariants

Figure 3: The first two affine moment invariants plotted against each other for 10 of the 20 subjects and their whitened values.

5 Conclusions

We proposed in this paper the use of the shape of the full spatio-temporal volume carved by a person who walks and the average image created from the spatio-temporal volume as biometrics. We showed

Table 1: Comparison of the experiment of [3] and our experiment

	The experiment of [3]	Our experiment
Classifier	The k-nearest neighbor rule	The support vector machine
Features	The gait signature consisting of 48 features	2D and 3D affine moment invariants consisting of 12 features
Classification rate [%]	92	94

(a) An average image (b) A differential image

Figure 4: An example of average images and differential images.

that affine moment invariants in conjunction with an SVM classifier may produce marginally better results than those based on gait analysis alone and k-nearest neighbour classification. The calculation of the affine moment invariants is very straight forward, unlike the individual cues extracted to characterize gait and certain characteristics of the human body. In addition, the affine moment invariants, being integrators, are more robust features than features based on differentiation.

One drawback may be that the spatio-temporal volumes we consider may be affected by clothing, so their use in unconstrained situations may not be robust. However, they may be used in conjunction with sensors that can "see through" clothing. Finally, another drawback is the use of the SVM classifier for multiple classes: it is less scalable than the k-nearest neighbour classifier. The use of k-nearest neighbour classifier (with k=1) in conjunction with 2D and 3D affine moment invariants yielded 90 % accuracy using whitened data. From these results, the proposed method is promising and merits further investigation.

References

[1] M. Nixon and J. Carter, "Automatic recognition by gait," *Proc. the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.

[2] Y. Sheikh and M. Shah, "Exploring the space of an action for human action recognition," *Proc. 10th IEEE Int. Conf. Computer Vision*, pp. 15–21, 2005.

[3] I. Bouchrika and M. Nixon, "People detection and recognition using gait for automated visual surveillance," *Proc. IEE Inter. Symp. Imaging for Crime Detection and Prevention*, 2006.

[4] D. Cunado, M. Nixon, and J. Carter, "Automatic extraction and description of human gait models for recognition purposes," *CVIU*, vol. 90, no. 1, pp. 1–41, 2003.

[5] C. Yam, M. Nixon, and J. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, 2004.

[6] R. Urtasun and P. Fua, "3d tracking for gait characterization and recognition," *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 17–22, 2004.

[7] L. Lee and W. Grimson, "Gait analysis for recognition and classification," *Proc. IEEE Conf. Face and Gesture Recognition*, pp. 155–161, 2002.

[8] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," *Proc. IEEE Conf. Face and Gesture Recognition*, pp. 372–377, 2002.

[9] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognition Letters*, vol. 17, no. 2, pp. 155–162, 1996.

[10] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The humanid gait challenge problem: data sets, performance, and analysis," *IEEE Trans. PAMI*, vol. 27, no. 2, pp. 162–177, 2005.

[11] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," *IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 351–356, 2002.

[12] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 863–876, 2006.

[13] S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in xyt," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 469–474, 1994.

[14] Y. Liu, R. Collins, and Y. Tsin, "Gait sequence analysis using frieze patterns," *Proc. the 7th European Conf. Computer Vision (ECCV'02)*, 2002.

[15] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "Eigengait: Motion-based recognition of people using image self-similarity," *3rd Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001.

[16] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. PAMI*, vol. 28, no. 2, pp. 316–322, 2006.

[17] J. Little and J. Boyd, "Recognizing people by their gait: The shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.

[18] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, 1993.

[19] A. G. Mamistvalov, "n-dimensional moment invariants and conceptual mathematical theory of recognition n-dimensional solids," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 819–831, 1998.

[20] D. Xu and H. Li, "3-d affine moment invariants generated by geometric primitives," *Proc. 18th IEEE Int. Conf. Pattern Recognition*, pp. 544–547, 2006.

[21] J. D. Shutler, M. G. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," *Proc. 4th Int. Conf. Recent Advances in Soft Computing*, pp. 66–71, 2002.

[22] L. Buturovic, "Pattern classification program," *http://pcp.sourceforge.net/*.