Gait Based Gender Classification Using Silhouette Image Gait Database

Sudheer Reddy Bandi[1], A Gopi Suresh[2] , Immanuel Alphonse[3]
[1,2,3]Assistant Professor, Dept of Information Technology, Tagore Engineering College, Chennai

Abstract

A realistic appearance-based representation to discriminate gender from side view gait sequence is introduced here. This gait representation is based on simple features such as moments extracted from orthogonal view video silhouettes of human walking motion. The silhouette image is divided into two regions. The first region includes from the head to the torso region whereas the second region is from torso to the feet region and for each region features are extracted. Finally we employ different pattern classifiers like KNN (K- Nearest Neighbor) and SVM (Support Vector Machine) to classify the gender. The division of two regions is based on the centroid of the silhouette image. The experimental results show that SVM classifier gives better results when compared to other classifiers. The classification results are expected to be more reliable than those reported in previous papers. The proposed system is evaluated using side view videos of CASIA dataset B.

Keywords: Appearance based features, binary moments, ellipse features, gait analysis, gender classification, and human silhouette.

1. INTRODUCTION

Recently, biometrics has increasingly attracted the attention as a key technology for realizing a more secure and safer society. Although most of the studies on biometrics focus on person authentication, namely hard bio metrics, it is also important to promote the recognition of properties such as gender, age and ethnicity, namely soft biometrics. The perception of gender recognition determines social interactions. Humans are very accurate at identifying gender from a face, voice and through gait. Among biometric modalities, gait has several promising properties such as availability at a distance from a camera even without the cooperation of the subject; hence gait based hard biometrics [13][14][15] has been extensively studied with the aim of realizing wide area surveillance and assistance with criminal investigation. Furthermore, gait based soft biometrics are also an active research area (e.g., gender classification [7][9][10], age group classification [16][17]).

There has been much work to classify gender from human faces. In early 1990s, various neural network techniques were employed for gender classification from a frontal face. In paper [1] Golomb et al. trained a fully connected two-layer neural network, SEXNET, to identify gender from face images. Brunelli and Poggio [2] developed HyperBF networks for gender classification in which two competing networks, one for male and the other for female, are trained using 16 geometric features. To sum up, some of these techniques are appearance-based methods and others are based on geometric features. In Moghaddam and Yang's paper [3], nonlinear SVM was investigated in gender classification for low-resolution thumbnail face (21-by-12 pixels) on 1,755 images from the FERET database.

Proceeding toward gender classification in emotional speech; in [4] Harb et al. proposed a method in which a set of acoustic and pitch features are used for gender identification. Concerning previous work on non-emotional speech, the system proposed by Zeng et al. [5] is based on Gaussian mixture models (GMMs) of pitch and spectral perceptual linear predictive coefficients.

Nevertheless, when compared to face or voice, gait can be perceived at a distance. In addition to these, gait has various advantages like non-contact, non-invasive and in general, does not require subject's willingness. This particular issue has stirred up the interest of the computer vision community in creating gait based gender recognition systems. A number of applications can be benefitted from the development of such systems. For example demographic analysis of systems, access control, biometric systems that use gender recognition to reduce the search space to half, etc.

Most of gait research has been addressed to biometric identification, which consists of predicting the identity of a person according to his/her way of walk. However, few recent works have used gait analysis for other classification tasks, such as gender recognition or age estimation. Regardless of the purposes two different approaches to describe gait can be considered. Some works extract dynamic features from subject's movements [12, 10], while others take static attributes related with the appearance of the subject [6, 7, 9]; what implicitly might contain motion information.

There have been a number of appearance based methods to classify gender from gait. Lee and Grimson analyzed the motion of 7 different regions of a silhouette [9]. Features for discrimination of gender were selected using Analysis of Variation (ANOVA), and an SVM was trained to categorize gender.

Similar to [9] Huang and Wang [7] also use ellipse features for gender classification. The difference is that Huang and Wang combine multiview features to improve the performance. Unlike Lee's and Huang's methods, in [12], the gait signature was represented as a sequential set of 2-D stick figures. Each 2-D stick figure is extracted from a human silhouette and contains head, neck, shoulder, waist, pelvis, knees, and ankles. Davis et al. presented a three-mode expressive-feature model for recognizing gender from point-light displays of walking people [10]. They developed a trimodal nature of walkers (posture, time, and gender) in which an efficient three-mode PCA representation was employed.

The previously introduced methods are based on human dynamic movement or silhouettes, and the background and clothing texture are all removed. Different from these methods, Cao et al. proposed a new method which recognizes gender from static full body images [8]. They use the Histogram of Oriented Gradients (HOG) as the feature, and Adaboost and Random Forest algorithms as classifiers.

Head and hair, back, chest and thigh are more discriminative for gender classification than other components. Body sway, waist-hip ratio, and shoulder-hip ratio are also indicative of a walker's gender. Males tend to swing their shoulders more than their hips, and females tend to swing their hips more than their shoulders. Males normally have wider shoulders than females, and females normally have thin waists and wider hips. Hair style and chest are two important body components for gender classification. However, the hair component is divided into head component and trunk component.

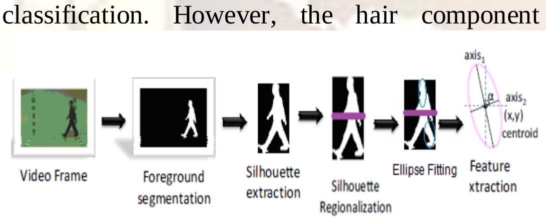

Fig 1. Process of Feature Extraction

Chest sometimes can be divided into the arm component (for fat persons), and sometimes into the trunk component (for slim persons). To study the effectiveness of different components, it is better to set the segmentation borders in the areas where both males and females are the same, such as the centre of the trunk. The upper body silhouettes convey more static information than the lower body silhouettes.

The rest of the paper is organized as follows. Methodology of the proposed work is given in Section 2. Performance analysis and results are explained in Section 3. Section 4 concludes along with some future works.

2. METHODOLOGY OF PROPOSED WORK

The approach proposed here for human gait characterization is based on the method introduced by Lee and Grimson [9], where the silhouette appearance of a side-view human walking sequence was described by static features that were further used for gender recognition. The architectural block diagram for proposed methodology is given in Fig 2. The work proposed here can be defined in terms of the following steps.

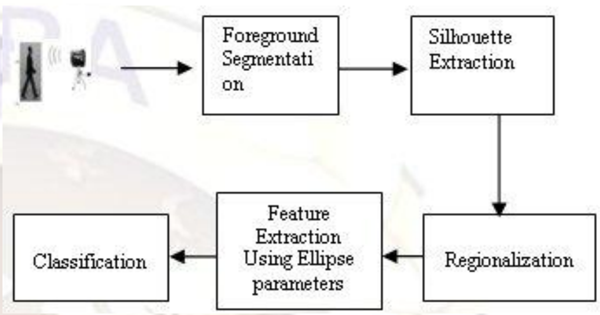

Fig 2. Architectural block diagram of proposed system.

2.1 Foreground segmentation

For each frame, the foreground (silhouette) is segmented from the background. A new binary frame with the silhouette highlighted is obtained. The algorithm used for foreground segmentation is frame difference approach.

2.2 Silhouette Extraction

In order to be robust to changes of clothing and illumination it is reasonable to consider the binary silhouette of the object, which is nothing but outline of the moving object with featureless interior. A reduced image defined by the bounding box that encloses the silhouette is extracted to normalize its size and location.

2.3 Regionalization

The bounding box is divided into two regions obtained from the silhouette centroid and from fixed percentages of the box sizes.

2.4 Feature Extraction

For each silhouette of a gait video sequence, we find the centroid and proportionally divide the silhouette into two regions as shown in Fig .1. The frontal-parallel view of the silhouette is divided into the front and back sections by a vertical line at the silhouette centroid. The parts above and below the centroid are equally divided in the horizontal direction, resulting in two regions that roughly correspond to: r1, head to torso region; r2, torso to feet region.

For each and every region ellipse is formed and its parameters are considered. The fitting of an ellipse region involves computing the mean and covariance matrix for the foreground pixels in the region. Since images contain a great deal of information, issues of representation become important. We show that the geometric properties of interest can be computed from projections of binary images. Binary images are used in many applications since they are the simplest to process, but they are such an impoverished representation of the image information that their use is not always possible. However, they are useful where all the information we need can be provided by the silhouette of the object and when we can obtain the silhouette of that object easily.

The various binary image properties are given as 0^{th}, 1^{st} and 2^{nd} order moments. The area is given by 0^{th} moment of object. The centroid is given by 1^{st} order moments as (M_{10}, M_{01}) and 2^{nd} order moments gives the orientation of the object with the axis. Most of these notes are adapted from Horn's book on Robot Vision [11].

Let f(x, y) be the binary image in which we want to fit an ellipse. Assume that the foreground pixels are 1 and the background pixels are 0 then the mean x and y of the foreground pixels or the centroid of the region is given by,

$$xmean=1/N*(\sum f(x,y)*x) \quad (1)$$

$$ymean=1/N*(\sum f(x,y)*y) \quad (2)$$

where N is the total number of foreground pixels:

$$N=\sum I(x, y) \quad (3)$$

The binary moments of image are given as follows

$$M_{pq}=\sum\sum x^p y^q f(x,y)$$
$$Area==\sum\sum f(x,y)$$
$$\mu_{00}=M_{00}$$
$$\mu_{01}= \mu_{10}=0$$
$$\mu_{11}= M_{11}\text{-}xmean* M_{01}= M_{11}\text{-}ymean* M_{01}$$
$$\mu_{20}=M_{20}\text{-}xmean*M_{10}$$
$$\mu_{02}= M_{02}\text{-}ymean*M_{01}$$
$$\mu'_{02}= \mu_{02}/ \mu_{00}= M_{02}/ M_{00}\text{-}ymean^2$$
$$\mu'_{20}= \mu_{20}/ \mu_{00}= M_{20}/ M_{00}\text{-}xmean^2$$
$$\mu'_{11}= \mu_{11}/ \mu_{00}= M_{11}/ M_{00}\text{-}xmean*ymean$$
$$\lambda_i= (\mu'_{20}+ \mu'_{02}\pm sqrt(4\mu_{11}{}^2+(\mu'_{20}\text{-} \mu'_{02})^2))/2$$
$$Major\ axis=4*sqrt(\lambda_{max})$$
$$Minor\ axis=4*sqrt(\lambda_{min})$$

Fig 3. Binary moments of an image.

From the binary moments the aspect ratio and angle of orientation are given as follows.

$$Aspect\ ratio=major\ axis/minor\ axis \quad (4)$$

$$Angle=0.5atan\ (2\ \mu'_{11}/(\mu'_{20}\text{-}\ \mu'_{02})) \quad (5)$$

The orientation is defined only modulo π so it is chosen to lie in a range of π appropriate for each region of the silhouette. The ellipse parameters extracted from each region of the silhouette are the centroid, the aspect ratio(l) of the major axis to minor axis, the angle of orientation (α) of major axis to x axis which forms the region feature vector $f(r_i)$ as in ("6,")

$$f(r_i)=(x_i,y_i,l_i,\alpha_i) \quad (6)$$

The features extracted from each frame of a walking sequence consists of features from each of the two regions, so the frame feature vector F_j of the j^{th} frame is given as

$$F_j=(f(r_1),f(r_2)) \quad (7)$$

In addition to these 8 features, we use one additional feature, *h*, the height (relative to body length) of the centroid of the whole silhouette to describe the proportions of the torso and legs.

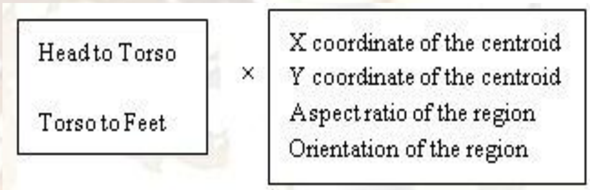

Fig 4. Representation of 8 features of a silhouette frame.

Given the region features across a gait sequence, we need a concise representation across time. We compute two types of features across time: the mean and standard deviation of region features across time. Specifically, the gait average appearance feature vector of a sequence *s* is,

$$s= (mean_j(F_j), std_j(F_j)) \quad (8)$$

Where *j* = 1. . . last frame and *s* is 8-dimensional. This feature set is very simple to compute and robust to noisy foreground silhouettes. Intuitively, the mean features describe

the average–looking ellipses for each of two regions of the body; taken together, the two ellipses describe the average shape of the body. The standard deviation features roughly describe the changes in the shape of each region caused by the motion of the body, where the amount of change is affected by factors such as how much one swings one's arms and legs.

However, the mean and standard deviation representation has the additional advantage that they are not seriously affected by the occasional frame skip that a foreground segmentation algorithm may undergo. Hence, the feature vector we compute for each gait sequence consists of the mean and standard deviations across time of each of the parameters for each region. The feature vector for one silhouette image is given in Table I.

TABLE I
FEATURE VALUES FOR SINGLE SILHOUETTE IMAGE

P \ R	X	Y	Aspect ratio (l)	Angle of orientation (α)
r1	11.6474	35.5723	11.9789	-3.0613
r2	19.6896	34.5848	5.6745	-2.9831

Where P=Parameters, R=Regions X= x coordinate of the centroid, Y= y coordinate of the centroid

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm the experiments are conducted in the CASIA database B and various classifiers are used to compare the results which are explained in the subsections given below.

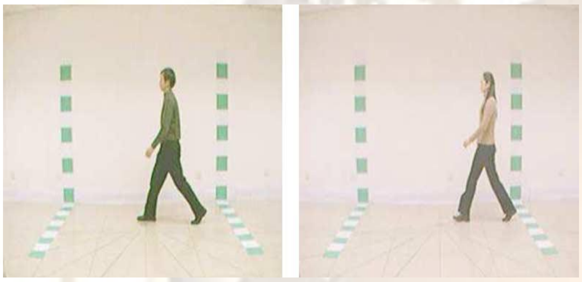

Fig 5. Male (left) and female (right) images from the CASIA Database.

3.1 Data Acquisition

In any pattern recognition system, the database used for evaluation is necessary. We choose the CASIA Gait Database (Dataset B) for our experiments as shown in Fig 5. There are 124 subjects (93 males and 31 females) in the CASIA Database (Dataset B), and most subjects are Asians. All the 31 females and 31 randomly selected males are used in the experiments. There are six video sequences for each subjects; altogether there are 372 sequences. All selected data were captured from the side view with normal clothes and without any bag as shown in Fig. 5.

3.2 Classification Results and Analysis

For each side view videos of CASIA gait database, we first generate silhouette images using background subtraction algorithm and then silhouette image features are extracted as explained in above section. Then we trained two classification models SVM and kNN with the 8 features and the genders are classified by the trained model set. We report results in terms of cumulative match scores. To calculate these scores, we conduct multiple tests using multiple test sequences. Each test sequence is compared to the sequences in the reference data base. Train data and test data is formed from the CASIA database B.

We used an implementation of support-vector machine and experimented with the linear, Multilayer Perceptron, Polynomial, Quadratic and Radial basis function kernels. The SVM's are trained using the 8 features of a single video and under the random-person vs. random-sequence conditions. The results for these tests conditions are listed in Table II. These features are also trained using KNN classifier. Four normal forms of walking are used to test the data. The average performance of all the data is taken and their results are tabulated in Table III.

The precision rate of SVM and kNN models are summarized in Table IV, from which it can be seen that the recognition performance based on SVM classifier is better than kNN. Precision rate of the classifiers are calculated using "(9,)".

$$\text{Precision}=\frac{No.of\ correctly\ classified\ test\ data}{Total\ No.of\ test\ data}\times 100 \quad (9)$$

TABLE III
CLASSIFICATION RATES OF DIFFERENT SVM KERNEL TYPES

SVM kernel Type	Recognition Rate
Linear	93.33%
Polynomial	81.67%
Quadratic	80.00%
Radial basis function	75.83%

In this work apart from recognition rate, other measures such as True Positive rate (TPr), True Negative rate (TNr), Geometric Mean and Area Under Classification (AUC) which are more appropriate for imbalanced problems are also considered and the values are tabulated in Table IV. The experiments for the proposed approach were conducted on a personal computer with an Intel Core 2 Duo processor (2.19 GHz) and 1 GB RAM configured with Microsoft Windows XP and Matlab 7.5 software with image processing toolbox and bio informatics toolbox. The comparison chart of all the specified performance measures is given in Fig. 6. The formulae for the measures are given below.

$$\text{TPr}=\frac{No.of\ correctly\ classified\ Positive\ data}{Total\ No.of\ Positive\ test\ data}\times 100 \quad (10)$$

$$\text{TNr}=\frac{No.of\ correctly\ classified\ Negative\ data}{Total\ No.of\ Negative\ test\ data}\times 100 \quad (11)$$

$$\text{GM}=\sqrt{TPr \times TNr} \quad (12)$$

AUC=(TPr+TNr)/2 (13)

TABLE IV
RECOGNITION RATES OF DIFFERENT CLASSIFIERS

Classifier Type	Recognition Rate or Precision rate
SVM	93.33%
KNN	83.33%

TABLE V
COMPARISON OF DIFFERENT PERFORMANCE MEASURES

Performance Measures	SVM	KNN
TPr	96.43%	86.67%
TNr	90.00%	80.00%
GM	93.27%	83.27%
AUC	93.34%	83.34%

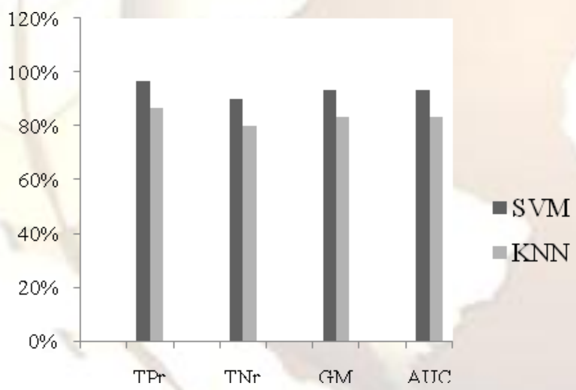

Fig. 6 Comparison Charts of Performance Measures

4. CONCLUSION

Gait-based gender classification is a new and interesting topic. This paper introduces a realistic appearance based representation of gait sequences for automatic gender recognition. An exhaustive study designed to evaluate the capacity of the CASIA Gait Database for gender recognition tasks was carried out. This dataset contains gait samples from 124 subjects, distributed in an unbalanced way with 31 women and 93 men. Our representation is rich enough to show promising results in these tasks. Various pattern classifiers like KNN and SVM are used for classification and the SVM gives better recognition results. This view and appearance dependent model of gait can be further extended to accommodate a multiple appearance model of a person and in conjunction with other recognition modalities.

REFERENCES

[1] B. A. Golomb, D. T. Lawrence and T. J. Sejnowski. "Sexnet: A neural network identifies sex from human faces," *In International Conference on Advances in Neural Information Processing Systems*, 1991.
[2] R. Brunelli and T. Poggio, "Hyperbf networks for gender classification," *In DRAPA Image Understanding Workshop*, 1992.
[3] B. Moghaddam and M. Yang. "Learning Gender with Support Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):707–711, May 2002.
[4] H. Harb and L. Chen, "Voice-based gender identification in multimedia applications," J. *Intelligent Information Systems*, 24(2):179–198, 2005.
[5] Y. Zeng, Z. Wu, T. Falk, and W. Y. Chan (2006). "Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech," *In Proc. 5th. IEEE Int. Conf Machine Learning and Cybernetics*, pages 3376–3379. China.
[6] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 145–155, 2008.
[7] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," in *Proc. 8th Asian Conf. Computer Vision*, 2007, pp. 462–471.
[8] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 725–728.
[9] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, May 2002, pp. 155–162.
[10] J. W. Davis and H. Gao, "Gender recognition from walking movements using adaptive three-mode PCA," in *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, Washington, DC, 2004, vol. 1.
[11] Berthold Klaus Paul Horn, "Robot Vision," *MIT Press*, 1986.

[12] J.-H.Yoo, D.Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," in *Proc. Advanced Concepts for Intelligent Vision Systems*, Antwerp, Belgium, Sep. 2005, pp. 138–145.

[13] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. International Series on Biometrics. Springer-Verlag, Dec. 2005.

[14] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[15] J. Han and B. Bhanu. Individual recognition using gait energy image. *Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316– 322, 2006.

[16] R. Begg. Support vector machines for automated gait classification. *IEEE Trans. on Biomedical Engineering*, 52(5):828–838, May 2005.

[17] J. Davis. Visual categorization of children and adult walking styles. In *Proc. of the Int. Conf. on Audio- and Video-based Biometric Person Authentication*, pages 295–300, Jun. 2001.