Learning Gender from Human Gaits and Faces

Caifeng Shan, Shaogang Gong and Peter W. McOwan
Department of Computer Science, Queen Mary, University of London, London E1 4NS, UK
{cfshan, sgg, pmco}@dcs.qmul.ac.uk

Abstract

Computer vision based gender classification is an important component in visual surveillance systems. In this paper, we investigate gender classification from human gaits in image sequences, a relatively understudied problem. Moreover, we propose to fuse gait and face for improved gender discrimination. We exploit Canonical Correlation Analysis (CCA), a powerful tool that is well suited for relating two sets of measurements, to fuse the two modalities at the feature level. Experiments demonstrate that our multimodal gender recognition system achieves the superior recognition performance of 97.2% in large datasets.

1. Introduction

Gender classification is an important visual task for human beings, as many social interactions critically depend on the correct gender perception. As visual surveillance and human-computer interaction technologies evolves, computer vision systems for gender classification will play an increasing important role in our lives, e.g., collecting valuable demographic information in a social environment.

As human faces provide important visual information for gender perception, a large number of studies have investigated gender classification from human faces [1, 2]. However, in the real-world unconstrained situations, due to the arbitrary walking direction and continuously varying head pose, face information sometimes is unreliable or unavailable. More crucially, with people walking at a distance, face information can not be measured reliably at low resolution. In these situations, human gait, or the style of walking, can provide important alternative cues for gender classification, as gaits can be detected and measured from arbitrary views and at a distance. Psychophysical studies [3, 4, 5] have shown that, even for point-light displays (filmed by attaching small point lights to the main joints of human body in a homogeneously dark background), people can recognize the gender of walkers. However, given the ability of humans to classify gender by the gaits, there have been few computer vision systems developed for gender recognition from gaits. This problem is relatively understudied, compared to facial gender classification.

In this paper, we investigate gender classification from human gaits in image sequences using machine learning methods. Considering each modality, face or gait, in isolation has its inherent weakness and limitations, we further propose to fuse gait and face for improved gender discrimination. We exploit Canonical Correlation Analysis (CCA), a powerful tool that is well suited for relating two sets of signals, to fuse the two modalities at the feature level. Experiments on large dataset demonstrate that our multimodal gender recognition system achieves the superior recognition performance of 97.2%. We plot in Figure 1 the flow chart of our multimodal gender recognition system.

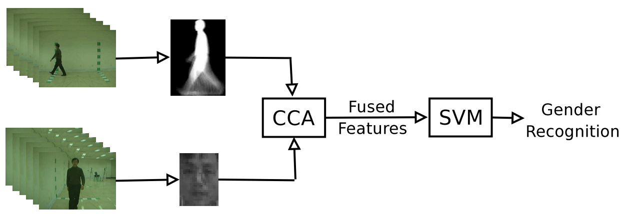

Figure 1: The flow chart of our multimodal gender recognition system.

2. Previous work

Most of the existing work on gender classification attempt to classify gender from human faces. In the early 1990s various neural network techniques were employed for gender classification from a frontal face [6, 7]. Some of these techniques are appearance-based methods, while others are based on geometrical features. Recently Moghaddam and Yang [1] investigated nonlinear Support Vector Machines (SVMs) for gender classification with low-resolution thumbnail face, and demonstrated the superior performance of SVMs to other classifiers. Walawalkar et al. [8] adopted SVMs for gender classification using audio and visual cues. Shakhnarovich *et al.*[2] developed a real-time face detection and demographic analysis (female/male and asian/non-asian) system using Adaboost, which delivers slightly better performance than the SVMs [1] on unaligned faces from real-world unconstrained video sequences.

Gender classification from human gaits has mainly been studied in the psychological field in last several decades [3, 4, 5]. Using point-light displays, not only the structural differences between male and female walkers, such as the shoulder-hip ratio and center-of-moment features of the torso [4], but also the temporal dynamic factors, such as arm swing and shoulder sway [5], have been examined. Many of these studies have focused on the manual identification of key features that enable the perceptual classification between female and male walking styles. However, to date there is no conclusive evidence as to which features actually drive the discrimination process. It seems that gender information is not a matter of a single feature, but rather involves multiple combined features. Troje [9] recently treated the analysis of biological motion as a linear pattern recognition problem, presenting a two-stage Principal Component Analysis (PCA) framework for recognizing gender. Davis and Gao [10] more recently presented an approach for gender recognition using an expressive three-mode PCA model.

The aforementioned studies on gender discrimination from gaits all used point-light displays from the aspect of biological motion. Lee and Grimson [11] adopted a computer vision algorithm to extract dynamics features of gaits from image sequences for gender classification. They divided the binary silhouette into seven parts roughly corresponding to head/shoulder, arms/torso (front and back), thighs (front and back), and calve/feet (front and back), then extracted moment based features from each part. Using SVMs as classifiers, their approach achieved performance of 84.5% on a dataset of 24 subjects.

3. Gender Recognition from Gaits

3.1. Gait Representation

Lee and Grimson [11] only considered dynamic features for gender representation. In our work, we investigate structural features and dynamic features of gaits for gender recognition, by adopting Gait Energy Image (GEI) [12], a recently proposed spatio-temporal compact representation of gaits. GEI has been demonstrated to be effective for representing gaits in the human identification problem [12, 13].

Using background substraction techniques, the walking subjects can be extracted from the original image sequences to derive binary silhouette image sequences. To make the gait representation insensitive to the distance between the camera and the subject, we perform silhouette preprocessing procedure including size normalization and horizontal alignment [12]. Some examples of normalized and aligned silhouette images are shown in Figure 2. The entire human gait sequence can be divided into cycles as human walking repeats at a stable frequency. We detect the gait cycles using the method adopted in [14].

Given the preprocessed binary silhouette image $B_t(x, y)$ at time t in a sequence, the GEI is defined as follows:

Figure 2: Examples of normalized and aligned silhouette images. The rightmost image is the corresponding GEI.

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B_t(x, y) \quad (1)$$

where N is the number of frames in the complete cycle(s) of a silhouette sequence, t is the frame number of the sequence, and x and y are values in the 2D image coordinate (see Figure 2 for an example of GEI). GEI reflects shapes of silhouette and their changes over the gait cycle, and it is not sensitive to incidental silhouette errors in individual frames.

3.2. Gender Classification: SVMs

A previous successful technique for gender classification is SVM [1, 11, 8], so we adopt SVM classifiers here. SVM is an optimal discriminant method based on the Bayesian learning theory. For the cases where it is difficult to estimate the density model in high-dimensional space, the discriminant approach is preferable to the generative approach. SVM performs an implicit mapping of data into a higher dimensional feature space, and then finds a linear separating hyperplane with the maximal margin to separate data in this higher dimensional space.

Given a training set of labeled examples $\{(x_i, y_i), i = 1, \ldots, l\}$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$, a new test example x is classified by the following function:

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b) \quad (2)$$

where α_i are Lagrange multipliers of a dual optimization problem that describe the separating hyperplane, $K(\cdot, \cdot)$ is a kernel function, and b is the threshold parameter of the hyperplane. The training sample x_i with $\alpha_i > 0$ is called the *support vector*, and SVM finds the hyperplane that maximizes the distance between the support vectors and the hyperplane. Given a non-linear mapping Φ that embeds the input data into the high dimensional space, kernels have the form of $K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle$. SVM allows domain-specific selection of the kernel function. Though new kernels are being proposed, the most commonly used kernel functions are the linear, polynomial, and Radial Basis Function (RBF) kernels.

4. Fusing Gaits and Faces for Gender Recognition

Each modality, gait or face, has its inherent weakness and limitations. Fusing gait and face cues in image sequences is a potential way to accomplish effective gender discrimination. In this study, we further present a multimodal gender recognition system by fusing gaits and faces.

Recently several attempts [15, 16, 13] have been made to integrate face and gait cues for the human identification problem. Shakhnarovich and Darrell [15] rendered virtual views for frontal face recognition and side-view gait recognition, which were then combined for identification. Zhou and Bhanu [13] combined side face and gait cues for human identification. All these existing studies have focused on the decision-level fusion of face and gait, while the feature-level fusion is understudied. This is mainly because the two modalities may have incompatible feature sets and the relationship between the different feature spaces is unknown. Here we propose to fuse face and gait cues at the feature level using CCA. Our motivation is that, as face and gait are two sets of measurements for human gender, conceptually the two modalities are correlated, and their relationship can be established using CCA. CCA derives a semantic "gender" space, in which the gait features and face features are compatible and can be effectively fused.

4.1. Canonical Correlation Analysis

CCA is a statistical technique developed by H. Hotelling [17] for measuring linear relationships between two multidimensional variables. It finds pairs of base vectors (i.e., canonical factors) for two variables such that the correlations between the projections of the variables onto these canonical factors are mutually maximized. CCA has been applied to computer vision and pattern recognition problems [18, 19, 20]. Borga [18] adopted CCA to find corresponding points in stereo images. Melzer *et al.*[19] applied CCA to model the relation between an object's poses with raw brightness images for pose estimation. Harsoon *et al.*[20] learned a semantic representation to web images and their associated text using CCA.

Given two zero-mean random variables $\mathbf{x} \in R^m$ and $\mathbf{y} \in R^n$, CCA finds pairs of directions $\mathbf{w}_x$ and $\mathbf{w}_y$ that maximize the correlation between the projections $x = \mathbf{w}_x^T \mathbf{x}$ and $y = \mathbf{w}_y^T \mathbf{y}$. The projections x and y are called *canonical variates*. More formally, CCA maximizes the function:

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\mathbf{w}_x^T \mathbf{x}\mathbf{y}^T \mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T \mathbf{x}\mathbf{x}^T \mathbf{w}_x] E[\mathbf{w}_y^T \mathbf{y}\mathbf{y}^T \mathbf{w}_y]}} = \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}} \tag{3}$$

where $\mathbf{C}_{xx} \in R^{m \times m}$ and $\mathbf{C}_{yy} \in R^{n \times n}$ are the *within-set covariance matrices* of $\mathbf{x}$ and $\mathbf{y}$, respectively, while $\mathbf{C}_{xy} \in R^{m \times n}$ denotes their *between-sets covariance matrix*.

The maximization problem can be solved by setting the derivatives of Eqn. (3), with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$, equal to zero, resulting in the eigenvalue equations as:

$$\begin{cases} \mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{w}_x = \rho^2 \mathbf{w}_x \\ \mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\mathbf{w}_y = \rho^2 \mathbf{w}_y \end{cases} \tag{4}$$

where the eigenvalues ρ^2 are the squared canonical correlations between x and y. A number of at most $k = \min(m, n)$ canonical factor pairs $\langle \mathbf{w}_x^i, \mathbf{w}_y^i \rangle, i = 1, \ldots, k$ can be obtained. It can be shown that the canonical variates x_i and y_i (corresponding to $\mathbf{w}_x^i$ and $\mathbf{w}_y^i$) are maximally correlated and, at the same time, uncorrelated with the previous pairs x_j and $y_j, j = 1, \ldots, i-1$.

4.2. Feature Fusion of Gait and Face

Given $G = \{\mathbf{x} | \mathbf{x} \in R^m\}$ and $F = \{\mathbf{y} | \mathbf{y} \in R^n\}$, where $\mathbf{x}$ and $\mathbf{y}$ are the feature vectors extracted from gaits and faces respectively, we apply CCA to establish the relationship between $\mathbf{x}$ and $\mathbf{y}$. Suppose $\langle \mathbf{w}_x^i, \mathbf{w}_y^i \rangle, i = 1, \ldots, k$ are the canonical factors pairs obtained, we can use d ($1 \leq d \leq k$) factor pairs to represent the correlation information. With $\mathbf{W}_x = [\mathbf{w}_x^1, \ldots, \mathbf{w}_x^d]$ and $\mathbf{W}_y = [\mathbf{w}_y^1, \ldots, \mathbf{w}_y^d]$, we project the original feature vectors as $\mathbf{x}' = \mathbf{W}_x^T \mathbf{x} = [x_1, \ldots, x_d]^T$ and $\mathbf{y}' = \mathbf{W}_y^T \mathbf{y} = [y_1, \ldots, y_d]^T$ in the lower dimensional correlation space. We then combine the projected feature vector $\mathbf{x}'$ and $\mathbf{y}'$ to form the new feature vector as

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}' \\ \mathbf{y}' \end{pmatrix} = \begin{pmatrix} \mathbf{W}_x^T \mathbf{x} \\ \mathbf{W}_y^T \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_x & 0 \\ 0 & \mathbf{W}_y \end{pmatrix}^T \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \tag{5}$$

This fused feature vector effectively represents the multimodal information for gender discrimination.

5. Experiments

5.1. Data

We carried out experiments on the CASIA Gait Database (Dataset B) [21], currently one of the largest gait databases in the gait-research community. The database consists of 124 subjects aged between 20 and 30 years, of which 93 were male and 31 were female, and 123 were Asian and 1 was European. Each subject first walked naturally along a straight line six times, then put on his/her coat and walked twice, and finally walked twice carrying a bag (knapsack, satchel, or handbag). Each subject walked a total of ten times in the scene (6 normal + 2 with a coat + 2 with a bag). 11 cameras were uniformly set on the left hand side, with view angle interval of 18^o, so 11 video sequences from different views were captured simultaneously for every walking scenario (see Figure 3). There are a total of 13,640

(124×10×11) video sequences in the database, with 2 to 3 gait cycles in each sequence. The frame size is 320-by-240 pixel, and the frame rate is 25 fps.

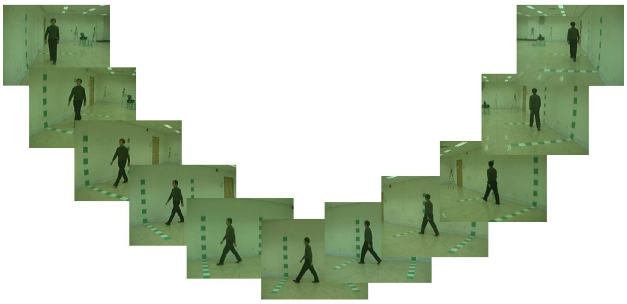

Figure 3: The walking sequences captured from 11 different views.

In our experiments we used video sequences from two views for gender recognition: frontal view for face cues and side view for gait cues. We selected video sequences of 119 subjects (88 Male and 31 Female) that suitable for gait and face analysis. In total 2,380 (119×10×2) video sequences were used in our experiments. Compared to the small dataset (24 subjects) used in the previous work [11], our study was performed on a much larger dataset.

As the database was collected for human gait analysis, there was no specific consideration of face data collection. Human faces were captured in an unconstrained environment like real-world surveillance scenario. The sequences contain facial expression changes, head pose variations, hair and glasses presented in the low-resolution faces. We first adopted a AdaBoost based face detector to detect face regions in each video sequence. Then, for simplicity, we manually labeled the three points (two eyes and the mouth) of the detected face with the best resolution in a sequence, and normalized the face as a 30-by-22 pixel thumbnail to represent the video sequence. That is, we extracted a face image for each video sequence. Video-based facial gender classification is a subject of our future research. To derive gait data, we computed the GEI for each video sequence. We show the processed face images and GEIs of 20 subjects (10 female + 10 male) in Figure 4, where the first row of GEIs are normal walking, and the second row is carrying a bag, while the bottom row is with wearing his/her coat.

5.2. Gender Recognition from Gaits

To evaluate the algorithms' generalization ability, we adopted a 5-fold cross-validation test scheme in all recognition experiments. That is, we divided the data set randomly into five groups with roughly equal (female and male) subjects, and then used the data from four groups for training and the left group for testing; the process was repeated five times for each group in turn to be tested. We report the average recognition rates (with the standard deviation) here. In all experiments, we set the soft margin C value of SVMs to infinity so that no training error was allowed. Meanwhile, each training and testing vector was scaled to be between -1 and 1. With regard to the hyper-parameter selection of Polynomial and RBF kernels, as suggested in [22], we carried out grid-search on the kernel parameters in the 5-fold cross-validation. The parameter setting producing the best cross-validation accuracy was picked. We used the SVM implementation in the publicly available machine learning library SPIDER [1] in our experiments.

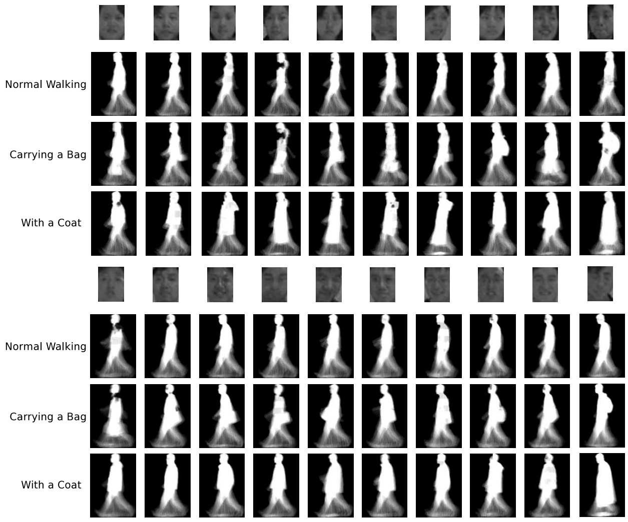

Figure 4: The extracted face images and GEIs of 20 subjects. (*Top*) Female; (*Bottom*) Male.

Classifier	Recognition Rates		
	Overall	Male	Female
SVM (Linear/Polynomial)	94.2±2.1%	97.5±3.2%	84.7±10.4%
SVM (RBF)	93.6±2.3%	96.8±3.9%	84.4±10.7%
PCA+LDA	94.5±1.9%	98.0±2.4%	84.6±9.6%

Table 1: Experimental results of gait-based gender recognition.

We report the results of gait-based gender recognition in Table 1. It is observed that GEIs based SVMs produce high overall recognition rates (93-94%), and the linear kernel and the (1st degree) polynomial kernel provide the same performance, slightly better than the RBF kernel. The number of support vectors of SVMs with different kernels were 13-16 percent of the total number of training samples. It is indicated that, for the GEI based gait representation, the linear decision surface is able to effectively classify gender, although there are many variations in GEIs due to wearing a coat or carrying a bag (as shown in Figure 4). To verify this, we further performed experiments with the linear subspace method PCA+LDA, which has frequently been used for the

[1] *http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html*

appearance-based object recognition. PCA reduces the dimension of feature space, and LDA identifies the most discriminant features. A nearest-neighbor classifier was used in our experiments. The experimental results summarized in Table 1 show that PCA+LDA achieves similar performance to the linear/polynomial kernels. Therefore, GEI is an effective gait representation for gender recognition, based on which the linear decision surface can discriminate gender with high confidence. The performance of GEI is also much better than that of dynamic features (84.5%) used in [11].

5.3. Gender Recognition from Gaits and Faces

Before fusing gait and face modalities, we first performed gender recognition with faces, and report the results in Table 2. By comparing Table 1 and Table 2, we can see that recognition results based on faces alone were consistently inferior to that based on gaits, which indicates that it is hard to learn human gender from low-resolution faces captured in unconstrained environments. For face-based gender recognition, SVMs have a clear margin of superiority over the linear subspace method PCA+LDA; the polynomial kernel also achieved the same performance with the linear kernel, but RBF kernel was found to perform best. The results we obtained reinforce the findings reported in [1]. This indicates that the face data can be better gender classified by the nonlinear decision surfaces. The number of support vectors of the linear/polynomial kernels were 23-24 percent of the total number of training samples, while the RBF kernel employed 25-39 percent. The SVMs' performance of 87-90% we obtained is inferior to that reported in [1]. This is because our face data was captured in an unconstrained real-world scenario, with the presence of facial expression changes, head pose variations, various hair styles and glasses, so it is more complex than the face images of FERET database used in [1].

Classifier	Recognition Rates		
	Overall	Male	Female
SVM (Linear/Polynomial)	87.5±1.8%	92.3±2.1%	74.3±10.3%
SVM (RBF)	90.4±1.8%	96.0±2.1%	74.6±9.7%
PCA+LDA	76.2±1.8%	79.6±3.5%	66.2±7.7%

Table 2: Experimental results of face-based gender recognition.

We then fused gait and face cues at the feature level using CCA. Different numbers of CCA factor pairs can be used to project the original gait and face feature vectors to a lower dimensional CCA feature space, and the recognition performance varies with the dimensionality of the projected CCA features. We first tested SVM (Linear) with the CCA features of different dimensions. We plot in Figure 5 the average recognition rates of SVM (Linear) versus CCA dimensionality reduction. It is observed that the projected CCA features of gaits and faces with 90-dimension provide the best performance. Hence we carried out subsequent experiments with CCA features of 90-dimension. For comparison, we also performed experiments using direct feature fusion, that is, concatenating the original gait and face feature vectors to derive a single feature vector. We report these experimental results in Table 3, where it shows the linear kernel also achieves the same performance as the polynomial kernel. We also plot bar graphs of recognition performance in Figure 6. We can see that the direct feature fusion produces only slight performance improvement over the single modality, while our proposed CCA feature fusion consistently provides the best recognition results. This is because CCA captures the relationship between the feature sets in different modalities, and the fused CCA features effectively represent information in each modality, removing noisy and redundant data. More crucially, the CCA feature fusion bring significant time and space benefit, for example, compared to the high dimensionality (4,160) in the direct feature fusion. the compact 180-dimension CCA features reduce the memory space by order 23. Another strength of the CCA feature fusion is that it always produces the smallest standard deviation of cross-validation, which demonstrate it is more robust than each single modality and the direct feature fusion. The performance 97.2% that the CCA feature fusion based SVM (RBF) obtained is better than 96.6% reported in [1], and, to our best knowledge, is the best gender recognition performance reported so far in the published literature.

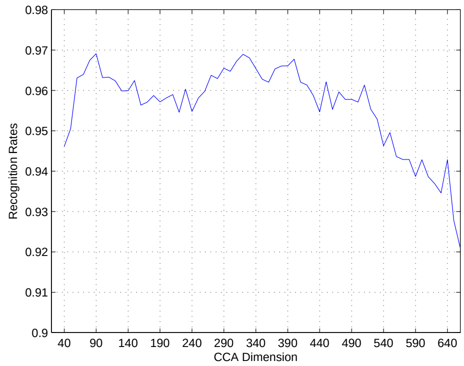

Figure 5: Recognition rates of SVM (Linear) versus dimensionality reduction of CCA.

We note that, in Table 1 - 3, all the female recognition rates are poorer than the male (with larger variance). In previous studies [1, 2], different classifiers also had higher error rates in classifying females. This phenomenon is possibly because the female gaits and faces have less prominent and distinct features, for example, the female has much variation in their hair styles and clothing. Another possible reason is the unbalance data set (88 Male and 31 Female). An encouraging observation is the female recognition performance based on each single modality is improved much by the CCA feature fusion (from 74-84% to 91-92%) which is significant.

	SVM (Linear/Polynomial)			SVM (RBF)			Feature Dimension
	Overall	Male	Female	Overall	Male	Female	
Direct Feature Fusion	95.6±1.7%	98.3±2.4%	88.0±8.6%	94.5±1.8%	97.4±3.1%	86.3±8.5%	4,160
CCA Feature Fusion	96.9±1.1%	99.0±1.1%	91.0±5.2%	97.2±0.8%	99.0±1.3%	92.0±4.6%	180

Table 3: Experimental results of gender recognition by fusing gaits and faces.

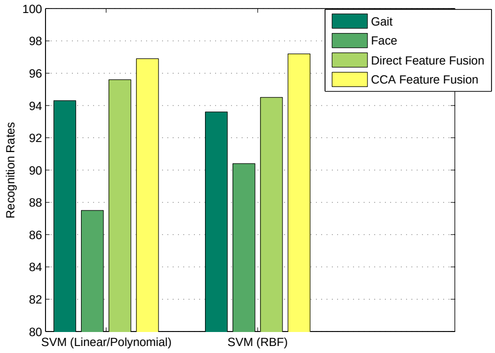

Figure 6: Gender recognition using different features.

6. Conclusions

In this paper, we investigate an important but understudied problem in visual surveillance, gender classification from human gaits. We also propose a method to effectively fuse gait and face at the feature level for improved gender discrimination. Experiments demonstrate that our multimodal gender recognition system achieves the superior recognition performance of 97.2% in large datasets.

References

[1] B. Moghaddam and M. Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, May 2002.

[2] G. Shakhnarovich, P. A. Viola, and B. Moghaddam. A unified learning framework for real time face detection and classification. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2002.

[3] L. T. Kozlowski and J. E. Cutting. Recognizing the sex of a walker from dynamic point-light display. *Perception & Psychophysics*, 21(6):575–580, 1977.

[4] C. D. Barclay, J. E. Cutting, and L. T. Kozlowski. Temporal and spatial actors in gait perception that influence gender recognition. *Perception & Psychophysics*, 23(2), 1978.

[5] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society: Biological Sciences*, 258(1353), 1994.

[6] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems*, 1991.

[7] R. Brunelli and T. Poggio. Hyperbf networks for gender classification. In *DRAPA Image Understanding Workshop*, 1992.

[8] L. Walawalkar, M. Yeasin, A. Narasimhamurthy, and R. Sharma. Support vector learning for gender classification using audio and visual cues. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(3):417–439, 2003.

[9] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):371–387, 2002.

[10] J. W. Davis and H. Gao. An expressive three-mode principal components model for gender recognition. *Journal of Vision*, 4(5):362–377, 2004.

[11] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2002.

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

[13] X. Zhou and B. Bhanu. Integrating face and gait for human recognition. In *CVPR Workshop on Biometrics*, 2006.

[14] S. Sarkar, Phillips P. J., Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2), 2005.

[15] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2002.

[16] A. Kale, A. K. R. Chowdhury, and R. Chellappa. Fusion of gait and face for human identification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2004.

[17] H. Hotelling. Relations between two sets of variates. *Biometrika*, 8:321–377, 1936.

[18] M. Borga. *Learning Multidimensional Signal Processing*. PhD thesis, Linkoping University, SE-581 83 Linkoping, Sweden, 1998. Dissertation No 531.

[19] T. Melzer, M. Reiter, and H. Bischof. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 39(9):1961–1973, 2003.

[20] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis; an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[21] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *International Conference on Pattern Recognition (ICPR)*, 2006.

[22] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Taipei, 2003.