Support Vector Machines for Visual Gender Classification

Abstract

Support Vector Machines (SVMs) are investigated for visual gender classification with low-resolution "thumbnail" faces (21-by-12 pixels) processed from 1,755 images from the FERET face database. The performance of SVMs (3.4% error) is shown to be superior to traditional pattern classifiers (Linear, Quadratic, Fisher Linear Discriminant, Nearest-Neighbor) as well as more modern techniques such as Radial Basis Function (RBF) classifiers and large ensemble-RBF networks. Surprisingly, SVMs also out-performed human test subjects at the same task: in an experimental study involving 30 human test subjects ranging in age from mid-20s to mid-40s, the average error rate was 32% for the same "thumbnails" and 6.7% with high-resolution images (still nearly twice the error rate of SVMs). The difference between low and high-resolution inputs with SVMs was only 1% thus demonstrating a degree of robustness and relative scale invariance.

1 Introduction

This paper is concerned with the problem of classifying gender from thumbnail face images in which only the main facial regions appear, i.e., without hair information. The motivation for using such images is two fold. First, humans change their hair styles frequently. Therefore face images are usually cropped to keep only the main facial regions in a robust face recognition method[1]. It has also been shown that better recognition rates can be achieved for methods using hairless face images [10]. Second, we investigate the amount of face information required for a classifier to learn male and female patterns. Previous studies on gender classification either use large images with hair information or a small dataset for experiments. We show that SVM classifiers are able to learn and classify gender from a large set of thumbnail images with high accuracy.

Recently, SVMs have been successfully applied to key functional tasks in computational face-processing. These include face detection [13], face pose discrimination [9] and face recognition [15]. In this paper, we apply SVMs for gender classification using thumbnail images and compare their performance with traditional classifiers (Linear, Quadratic, Fisher Linear Discriminant, and Nearest Neighbor) and more modern techniques such as RBF networks and large ensemble-RBF classifiers. We also compare the performance of SVM classifiers with the performance of human test subjects on high and low resolution images.

Our approach to gender classification is illustrated in

[1]Our method aims to crop every face such that as little hair appears in an image as possible.

Figure 1. A face image is preprocessed using an automatic face-processing system for normalizing for translation, scale as well as slight rotations. The resulting output faceprints are standardized to 80-by-40 pixels, and further subsampled to 21-by-12 pixels for low resolution experiments. We then train and test each classifier with the face images using five fold cross validation. In particular, we show that limited information from a thumbnail image accounts for accurate gender classification in SVM classifiers.
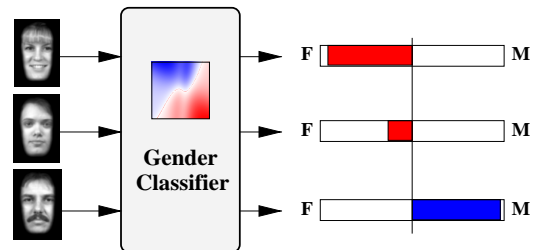

Figure 1. Gender classifier.

Although humans are extremely good at classifying gender from face images, our experiments show that most people have difficulty in classifying gender from hairless high resolution images. Furthermore human error rate in gender classification using low resolution images increases almost ten fold, while SVM classifiers show almost no difference in error rate. Note that no hair information has been used, in both machine and human gender classification experiments. This is in contrast to other experiments reported in literature where all but one method use hair information in gender classification.

This paper is organized as follows. We discuss related work in gender classification from the perspective of computational face processing in Section 2. Section 3 describes an automatic face processing system to locate and normalize face images into standard sizes. In Section 4, we describe the classifiers used in the experiments. Experimental results on these classifiers are presented in Section 5. We conclude this paper with some comments in Section 6.

2 Related Work

Questions regarding gender classification have been investigated from both psychological and computational perspective. Although gender classification has attracted much attention in psychological studies [1, 4, 6, 14], relatively few learning based vision methods have been proposed. In this section, we review only the methods in the latter category.

Gollomb, Lawrence and Sejnowski trained a fully connected two-layer neural network, SEXNET, to identify gender from 30-by-30 human face images [7]. Their experi-

ments on a set of 90 photos (45 males and 45 females) show an average error rate of 8.1% compared to an average error rate of 11.6% from a study of five human subjects. Cottrell and Metcalfe also applied neural networks for face emotion and gender recognition [5]. The dimensionality of a set of 160 64-by-64 face images (10 males and 10 females) is reduced from 4096 to 40 via an autoencoder network. These vectors are then given as inputs to another one layer network for training and recognition. Their experiments on gender classification report perfect results. Brunelli and Poggio [2] developed HyperBF networks for gender classification in which two competing RBF networks, one for male and the other one for female, are trained using 16 geometric features (e.g., pupil to eyebrow separation, eyebrow thickness, and nose width) as inputs. The results on a data set of 168 images (21 males and 21 females) show an average error rate of 21%. Similar to the methods by Golomb [7] and Cottrell [5], Tamura et al. [16] applied multilayer neural networks to classify gender from face images of multiple resolutions (from 32-by-32 to 16-by-16 and 8-by-8 pixels). Their experiments on 30 test images show that their network is able to determine gender from face images of 8-by-8 pixels with an average error rate of 7%. Instead of using a raster scan vector of gray levels to represent a face image, Wiskott et al. [18] used labeled graphs of two-dimensional views to describe faces. The nodes are labeled with jets which is a special class of local templates computed on the basis of wavelet transform, and the edges are labeled with distance vectors similar to geometric features in [3]. They use a small set of controlled model graphs of males and females to encode the general face knowledge. It represents the face image space and is used to generate graphs of new faces by elastic graph matching. For each new face, a composite face resembling the original one is constructed using the nodes in the model graphs. If the majority of the nodes in the composite graph are taken from female models, it is believed the face image have the same gender. The error rate of their experiments on a gallery of 112 face images is 9.8%. Recently Gutta, Wechsler and Phillips [8] propose a hybrid method which consists of ensemble of neural networks (RBFs) and inductive decision trees with Quinlan's C4.5 algorithm. Experimental results on a subset of face images of 384-by-256 pixels show that the best average error rate of their hybrid classifier is 4%.

3 Face Processing

In our study 256-by-384 FERET "mugshots" were preprocessed using an automatic face-processing system for normalizing for translation, scale as well as slight rotations. This system is described in detail in [11, 12] and uses maximum-likelihood estimation for face detection, affine warping for geometric shape alignment and contrast normalization for ambient lighting changes. The resulting out-
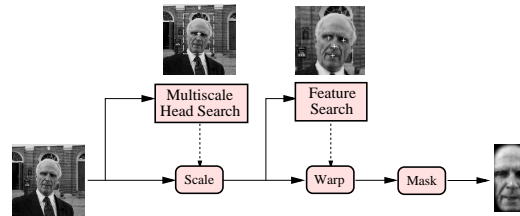

Figure 2. Face alignment system.

put "faceprints" seen in Figure 2 are standardized to 80-by-40 (full) resolution. These "faceprints" were further subsampled to 21-by-12 pixels for the low-resolution experiments. We processed 1,755 (1044 males and 711 females) FERET images for the experiments. Figure 3 shows some of the processed face images. Note that each processed face image contains as little hair information as possible.
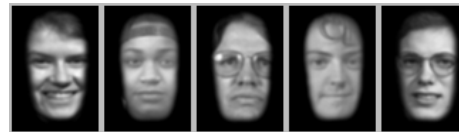

Figure 3. Some processed face images from FERET database.

4 Support Vector Machines

A Support Vector Machine is a learning algorithm for pattern recognition and regression problems [17]. One distinct characteristic of SVM is that it aims to find the optimal hyperplane such that the expected recognition error for the unseen test samples is minimized. According to the structural risk minimization inductive principle, a function that describes the training data well and belongs to a set of functions with lowest VC dimension will generalize well regardless of the dimensionality of the input space [17]. Based on this principle, the SVM adopts a systematic approach to find a linear function that belongs to a set of functions with lowest VC dimension. The SVM also provides non-linear function approximations by mapping the input vectors into a high dimensional feature space where a linear hyperplane can be constructed. Although there is no guarantee that a linear hyperplane will always exist in the high dimensional feature space, in practice it is quite possible to construct a linear SVM in the projected space.

Given a set of samples $(\mathbf{x}_1, y_1)$, $(\mathbf{x}_2, y_2)$, ..., $(\mathbf{x}_l, y_l)$ where $\mathbf{x}_i$ $(\mathbf{x}_i \in R^N)$ is the input vector of N dimension and y_i is its label $(y_i \in \{-1, 1\})$ for a recognition problem, SVM aims to find the optimal hyperplane that leaves the largest possible fraction of data points of the same class on

the same side while maximizing the distance of either class from the hyperplane (margin). Vapnik [17] shows that maximizing the margin distance is equivalent to minimizing the VC dimension in constructing an optimal hyperplane. The problem of finding the optimal hyperplane is thus posed as a constrained optimization problem and solved using quadratic programming techniques. The optimal hyperplane is in the form

$$f(\mathbf{x}) = \sum_{i=1}^{l} y_i \alpha_i \cdot k(\mathbf{x}, \mathbf{x}_i) + b$$

where $k(\cdot, \cdot)$ is a kernel function and the sign of $f(\mathbf{x})$ determines the label of $\mathbf{x}$. Constructing an optimal hyperplane is equivalent to determining nonzero α_i. Any vector $\mathbf{x}_i$ that corresponds to a nonzero α_i is a *supported vector* (SV) of the optimal hyperplane. One desirable feature of SVM is that the number of support vectors is usually small, thereby producing a compact classifier.

For a linear SVM, the kernel function is just the simple dot product of vectors in the input space while the kernel function in a nonlinear SVM projects the samples to a feature space of higher (possibly infinite) dimensions via a nonlinear mapping function:

$$\Phi : R^N \to F^M, \quad M \gg N$$

and construct a hyperplane in F. The motivation is that it is more likely to find a linear function, as done in linear SVM, in the high dimensional feature space. Using Mercer's theorem, the expensive calculations in projecting samples into high dimensional feature space can be reduced significantly by using a suitable function k such that

$$k(\mathbf{x}, \mathbf{x}_i) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)$$

where Φ is a nonlinear projection function. Several kernel functions, such as polynomial functions and radial basis functions, have been shown to satisfy Mercer's theorem and been used in nonlinear SVMs. By using different kernel functions, the SVM algorithm can construct a variety of learning machines, some of which coincide with classical architectures. However, this also results in a drawback since one needs to find the "right" kernel function in using nonlinear SVMs.

5 Experimental Results

We used a set of 1,755 (1044 males and 711 females) thumbnail images for experiments with five fold cross validation. The high and low resolution images, processed using methods described in the previous section, are 80-by-40 and 21-by-12 pixels respectively. Each image is represented by a raster scan eight-bit vector. The average size of training set is 1496 (793 males and 713 females) and average size of

test set is 259 (133 males and 126 females). All the experiments are carried out using low resolution images, unless otherwise specified, with five fold cross validation. After applying numerous kernels in SVM experiments, we noted that the SVM with RBF kernel performs best, followed by cubic polynomial kernel. In the large ensemble RBF experiment, the number of neurons is increased until the mean square error is below a target value. The average number of neurons in the large ensemble RBF is 1289. Our experiments show that the average number of neurons in the large ensemble RBF is 1289 (i.e., on the same order as the number of training examples), which indicates the inherent complexity of the gender classification problem. On the other hand, the number of neurons in classical RBF experiments was heuristically predetermined and set to be 20 prior to actual training and testing. Quadratic, linear and Fisher linear discriminant classifiers are also used in our experiments where two Gaussians are used to model the two distributions. Table 1 shows the experimental results on gender classification using thumbnail images.

Table 1. Experimental results using thumbnail images.

Classifier	Error Rate		
	Overall	Male	Female
SVM with RBF kernel	**3.38%**	**2.05%**	**4.79%**
SVM with cubic polynomial kernel	**4.88%**	**4.21%**	**5.59%**
Large Ensemble of RBF	5.54%	4.59%	6.55%
Classical RBF	7.79%	6.89%	8.75%
Quadratic classifier	10.63%	9.44%	11.88%
Fisher linear discriminant	13.03%	12.31%	13.78%
Nearest neighbor	27.16%	26.53%	28.04%
Linear classifier	58.95%	58.47%	59.45%

Both SVMs outperform other classifiers significantly. Although the performance of large ensemble of RBF is close to SVMs, the number of neurons used in the large ensemble of RBF is on the same order of the training set. On the other hand, the number of support faces (i.e., support vectors) is about 20% of the training set (for both SVMs). We have also applied SVMs to high resolution images for experiments. It is worth noting that both SVMs perform equally well in both low and high resolution experiments with only 1% error rate difference. This demonstrates a degree of robustness and relative scale invariance. Figure 4 shows a few support faces, both male and female, in one SVM classifier using RBF kernel. The sold line illustrates the optimal separating hyperplane and the dashed lines indicate the margins. Note that only few support faces are needed to construct a hyperplane for gender classification. In this diagram the support faces are arranged in accordance with their geometric interpretation, i.e., the pair of support faces are the closest pair of face images in the projected high dimensional space.
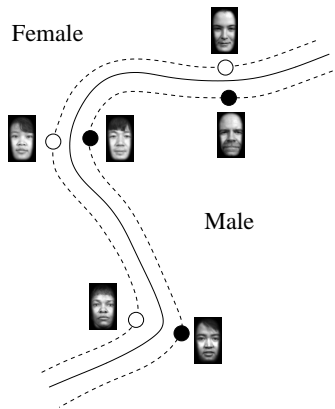
Figure 4. Support faces and the optimal separating hyperplane.

In order to calibrate the performance of SVM classifiers, human subject experiments on high and low resolution images have also been conducted. Thirty subjects (22 males and 8 females) ranging in age from mid-20s to mid-40s participated in the experiment with high resolution images and ten subjects (6 males and 4 females) perform in the experiment with low resolution images. All subjects are asked to classify the gender from the face images (presented in random order) as best as they can without time constraints. The experimental results are shown in Table 2. From Tables 1 and 2, it is clear that SVMs perform better than humans in both high resolution and low resolution experiments. The results also suggest that male and female images can be modeled by Gaussian distributions with the "right number" of clusters in a high dimensional space since large ensemble RBF networks perform equally well as human subjects while SVMs perform better than RBF networks. It is not surprising that human subjects perform far better with high resolution images than with low resolution images. On the other hand, SVMs perform almost equally well in both experiments. Note also that all the classifiers have higher error rates in classifying female images. This can be explained by the fact that there are more facial features in male images than female ones.

Table 2. Human subject experiments.

Gender of human subject	Error Rate	
	High resolution	Low resolution
Male	7.02%	30.87%
Female	5.22%	30.31%
Combined	**6.54%**	**30.65%**

6 Discussion and Conclusion

We have investigated SVMs for visual gender classification with low-resolution "thumbnail" faces (21-by-12 pix-

els) processed from 1,755 images from the FERET face database. The performance of SVMs (3.4% error) is shown to be superior to traditional pattern classifiers (Linear, Quadratic, Fisher Linear Discriminant, Nearest-Neighbor) as well as more modern techniques such as Radial Basis Function (RBF) classifiers and large ensemble-RBF networks. SVMs also out-performed human test subjects at the same task: in an experimental study involving 30 human test subjects ranging in age from mid-20s to mid-40s, the average error rate was 32% for the same "thumbnails" and 6.7% with high-resolution images (still nearly twice the error rate of SVMs). The difference between low and high-resolution inputs with SVMs was only 1% thus demonstrating a degree of robustness and relative scale invariance.

References

[1] V. Bruce, A. M. Burton, N. Dench, E. Hanna, P. Healey, O. Mason, A. Coombes, R. Fright, and A. Linney. Sex discrimination: How do we tell the difference between male and female faces? *Perception*, 22:131–152, 1993.

[2] R. Brunelli and T. Poggio. Hyperbf networks for gender classification. In *Proceedings of the DARPA Image Understanding Workshop*, pages 311–314, 1992.

[3] R. Brunelli and T. Poggio. Face recognition : Features vs. templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10), October 1993.

[4] A. M. Burton, V. Bruce, and N. Dench. What's the difference between men and women? evidence from facial measurement. *Perception*, 22:153–176, 1993.

[5] G. W. Cottrell. Empath: Face, emotion, and gender recognition using holons. In *Advances in Neural Information Processing Systems*, pages 564–571, 1991.

[6] B. Edelman, D. Valentin, and H. Abdi. Sex classification of face areas: how well can a linear neural network predict human performance. *Journal of Biological System*, 6(3):241–264, 1998.

[7] 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*, pages 572–577, 1991.

[8] S. Gutta, H. Wechsler, and P. J. Phillips. Gender and ethnic classification. In *Proceedings of the IEEE International Automatic Face and Gesture Recognition*, pages 194–199, 1998.

[9] J. Huang, X. Shao, and H. Wechsler. Face pose discrimination using support vector machines. In *Proc. of 14th Int'l Conf. on Pattern Recognition (ICPR'98)*, pages 154–156, August 1998.

[10] H.-Y. M. Liao, C.-C. Han, G.-J. Yu, H.-R. Tyan, M. C. Chen, and L.-H. Chen. Face recognition using a face-only database: A new approach. In *Proceedings of Asian Conference on Computer Vision*, volume 1352 of *Lecture Notes in Computer Science*, pages 742–749. Springer, 1998.

[11] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE Proceedings of the Fifth International Conference on Computer Vision (ICCV'95)*, Cambridge, USA, June 1995.

[12] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-19(7):696–710, July 1997.

[13] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136, 1997.

[14] A. J. O'Toole, T. Vetter, N. F. Troje, and H. H. Bulthoff. Sex classification is better with three-dimensional structure than with image intensity information. *Perception*, 26:75–84, 1997.

[15] P. J. Phillips. Support vector machines applied to face recognition. In M. S. Kearns, S. Solla, and D. Cohen, editors, *Advances in Neural Information Processing Systems 11*, volume 11, pages 803–809. MIT Press, 1998.

[16] S. Tamura, H. Kawai, and H. Mitsumoto. Male/female identification from 8 × 6 very low resolution face images by neural network. *Pattern Recognition*, 29(2):331–335, 1996.

[17] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[18] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition and gender determination. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 92–97, 1995.