

Gender Classification of Face Images: The Role of Global and Feature-Based Information

Samarasena Buchala¹, Neil Davey¹, Ray J.Frank¹, Tim M.Gale^{1,2},
Martin J.Loomes¹, Wanida Kanargard¹

¹ Department of Computer Science, University of Hertfordshire, College Lane, Hatfield,
AL10 9AB, UK

{S.Buchala, N.Davey, R.J.Frank, T.Gale, M.J.Loomes,
W.Pensuwon}@herts.ac.uk

² Department of Psychiatry, QEII Hospital, Welwyn Garden City, AL7 4HQ, UK

Abstract. Most computational models of gender classification use global information (the full face image) giving equal weight to the whole face area irrespective of the importance of the internal features. Here we use a two-way representation of face images that includes both global and featural information. We use dimensionality reduction techniques and a support vector machine classifier and show that this method performs better than either global or feature based representations alone.

1 Introduction

Most computational models of gender classification use whole face images, giving equal weight to all areas of the face, irrespective of the importance of internal features for this task. In this paper we evaluate the importance of global and local features based on experimentation. Global processing largely deals with coarse information like shape and configuration of internal features, while featural processing involves more detailed representation (e.g. eyes, mouth etc). We use these two representations and use a Support Vector Machine (SVM) classifier for gender classification. As the face images data have a very high dimensionality, we also use dimensionality reduction techniques before classification.

The remainder of the paper is organised as follows. Related work is discussed in the next Section. Section 3 discusses the methodology used for this study. Section 4 discusses the experimental results. We conclude with some discussion in Section 5.

2 Related work

The gender classification problem has attracted researchers from the fields of Psychology and Computer Science. While the research in Psychology [1], [2], [3] is in the context of human vision and identifying differing features in males and females, the Computer Science research [4], [5], [6], [7], [8] is mostly from the perspective of face recognition. The computational models range from using pixel-based information to

representations obtained from geometric measurements. Studies also vary as in the size of training sets used and in the type of features present or absent (for example, some studies use hair information while others do not). Nevertheless, most models, specifically the pixel-based, use whole face images, where all features carry equal weight. These can be termed as global models.

3 Methodology

We use a two way representation of face images which embodies both global and featural information. From a 128×128 face image three sub-images are obtained as illustrated in Fig.1. A 32×64 pixel strip pertaining to the eyes region, taking the mid-point between the two eyes as a reference point, and a 32×64 pixel strip pertaining to the mouth region, taking midpoint of the mouth as a reference point are extracted from each face image. These sub-images account for the featural information. The third sub-image is a 64×64 reduced resolution version of the original image and this represents global information. A similar type of face representation was also used by Luckman et al [9] for their computational model of familiar face recognition.

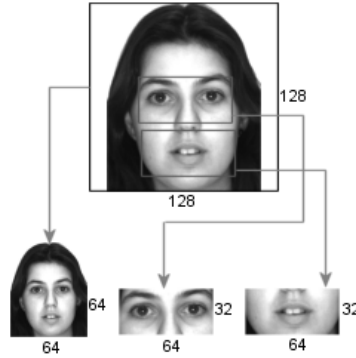


Fig. 1. Three sub-images are obtained from the original 128×128 image. A 32×64 image pertaining to the eye region and a 32×64 image pertaining to the mouth region are extracted from the original image. The third sub-image is a 64×64 reduced resolution version of the original image.

As the face images data have a very high dimensionality and due to “curse of dimensionality” [10], we apply dimensionality reduction techniques before applying an SVM for classification.

Principal Component Analysis (PCA) [11] is a popular dimensionality reduction technique and linearly transforms a D dimensional dataset X to a d dimensional dataset Y , without significant loss of information, where $d \leq D$.

Self Organising Map (SOM) [12] is a nonlinear method and learns a mapping from a D dimensional input space X to a d dimensional output space Y by using principles of Vector Quantization and Topological Mapping.

Curvilinear Component Analysis (CCA) [13], a recent technique, has the ability to reduce the dimensionality of strongly-nonlinear data. The output is a free space which takes the shape of the submanifold of the data. CCA minimizes the following error function:

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (d_{i,j}^X - d_{i,j}^Y)^2 F_{\lambda}(d_{i,j}^Y) \quad \forall j \neq i. \quad (1)$$

Where $d_{i,j}^X$ and $d_{i,j}^Y$ are the Euclidean distances between points i and j in the input space X and output space Y respectively. $F_{\lambda}(d_{i,j}^Y)$ is the neighbourhood function. The idea of CCA is to match distances in the input and output spaces. However, preservation of larger distances may not be possible in the case of nonlinear data, as a global unfolding of the manifold is required to reduce the dimension. In this case, it is important that at least local (smaller) distances should be preserved. For this reason CCA uses the neighbourhood function which ensures the condition of distance matching is satisfied for smaller distances while it is relaxed for larger distances. For details of the update rule, the reader is referred to [13].

The classification is performed using an SVM. The SVM [14] is a recently developed learning method, for pattern classification and regression. The basic idea of the SVM is to find the optimal hyperplane that has the maximal margin of separation between the classes, while having minimum classification errors.

Given a set of examples and their labels $\{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$ where $\mathbf{y}_i \in \{-1, 1\}$, the optimal hyperplane is given as:

$$f(X) = \sum_{i=1}^N \alpha_i y_i k(X, X_i) + b. \quad (2)$$

Constructing the optimal hyperplane is equivalent to finding α_i with nonzero values. The examples corresponding to the nonzero α_i are called support vectors. $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function, which implicitly maps the example data points into a high dimensional feature space, and takes inner product in that feature space. The potential benefit of a kernel function is that the data is more likely to be linearly separable in the high dimensional feature space, and also the actual mapping to the higher-dimensional space is never needed. We used an RBF kernel in our experiments.

4 Experiments

Experiments are carried out using 400 frontal face (200 females and 200 males) grey scale images. The faces are from the following databases: FERET [15], AR [16], and BioId [17]. Three sub-images, as explained in the previous Section, are extracted for each of the 400 faces. Histogram equalization is then applied on all three sub-images

to normalize for different lighting conditions. We use five-fold cross validation, with 320 faces (160 females and 160 males) for each training set and 80 faces (40 females and 40 males) for each test set, and report average classification rates using a SVM classifier, with RBF kernel. Before applying classification, dimensionality reduction techniques discussed in Section 3 are applied on the sub-images data. For PCA reduction we use the first few principal components, which account for 95% of the total variance of the data, and project the data onto these principal components. As CCA has the ability to reduce the dimensionality of strongly-nonlinear data, we use an Intrinsic Dimension¹ estimation technique, the Correlation Dimension [18] and reduce the data dimension to this Intrinsic Dimension. For SOM reduction, the subspace dimensionality is chosen as 64 (8×8 output grid) for the whole face and 36 (6×6 output grid) for eyes and mouth sub-images.

First we present classification results on the sub-images data. As shown in Table 1, all three sub-images produced high classification rates, indicating a surprisingly high amount of gender information in each of them. The figures in the parentheses indicate the subspace dimensionality. Classification is performed on the composite data, obtained by combining the data from the three sub-images. It can be seen from Table 2 that PCA performed marginally better than CCA and SOM. However, CCA uses far fewer variables (70) than PCA (759). For a comparison, we also report the classification rates of the data of the original 128×128 faces. It can be seen from Table 2 that the composite data, which includes both global and featural information, performed significantly better than the purely global model. It can be seen from Fig.2. that the composite data outperformed all other data representations.

Table 1. Average classification rates of the sub-images by SVM classifier. Figures in the parentheses are the number of variables obtained after dimensionality reduction

Feature	PCA	CCA	SOM
Eyes	85.5% (250)	82.75% (22)	80.25% (36)
Mouth	81.25% (253)	81.55% (22)	80.25% (36)
Full Face	87.5% (256)	87% (26)	83.25% (64)

Table 2. Classification rates of the composite image and original image data by SVM classifier. Figures in the parentheses are the number of variables obtained after dimensionality reduction

Feature	PCA	CCA	SOM
Composite	92.25% (759)	91.5% (70)	89.75% (136)
Original Full Face	86.5% (253)	85.5% (26)	83.25% (64)

¹ Due to correlations, linear and nonlinear, a D dimensional data may actually lie in a d dimensional space. This true dimension d is called Intrinsic Dimension, where $d \leq D$. As PCA accounts only linear correlations, it is unable to reduce the data dimension to its intrinsic dimension, when the correlations are nonlinear.

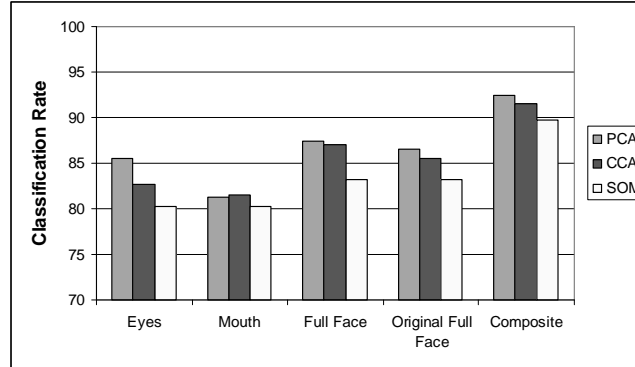


Fig. 2. Average classification rates on different features.

5 Discussion and Conclusion

Hair, especially for females, forms a major part of the image and has a dominating affect on the classification. Many males with long hair and females with short hair were misclassified when the original full face images are used. The global and feature based model largely solved this problem, by reducing the affect of misleading hair-styles, while not removing important hair information. Fig.3. shows examples of individual faces that are misclassified when the original full face images are used and classified correctly by the global and feature based model.

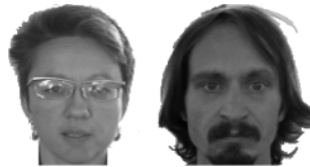


Fig. 3. Examples of the faces that are misclassified due to hair style of the individuals.

The global and feature based model for gender classification presented here performs significantly better than the global and featural models individually. This model allows inspection of facial data at various component levels and the results presented suggest that all components carry high levels of gender information. We believe that this type of representation also acts as a weighting factor of information, where highly variable discriminatory information (like hair) alone does not affect classification.

We also investigated three dimensionality reduction techniques. The Performance of CCA, a nonlinear technique, is comparable to PCA, while it uses far fewer variables than PCA.

References

1. Bruce, V., et al., "Sex discrimination: how do we tell the difference between male and female faces?," *Perception*, 1993. **22**: p. 131-152.
2. Burton, A.M., V. Bruce, and N. Dench, "What's the difference between men and women? Evidence from facial measurement.," *Perception*, 1993. **22**: p. 153-176.
3. Abdi, H., D. Valentin, B. Edelman, and J.A. O'Toole, "More about the difference between men and women: evidence from linear neural networks and the principal component approach.," *Perception*, 1995. **24**: p. 539-562.
4. Golomb, B., A., D. Lawrence, T., and T. Sejnowski, J., "Sexnet: A neural network identifies sex from human faces.," *Advances in Neural Information Processing Systems*, 1991. **3**: p. 572-577.
5. Brunelli, R. and T. Poggio. "HyperBF networks for gender classification,". in *DARPA Image Understanding Workshop*. 1992.
6. Tamura, S., H. Kawai, and H. Mitsumoto, "Male/female identification from 8×6 very low resolution images by neural network.," *Pattern Recognition*, 1996. **29**no.(2): p. 331-335.
7. Moghaddam B. and M.-H. Yang, "Gender classification with support vector machines.," Technical Report : TR-2000-01, Mitsubishi Electric Research Laboratory. January, 2000.
8. Sun, Z., X. Yuan, G. Bebis, and S. Louis, J. "Neural-Network-based gender classification using genetic search for eigen-feature selection,". in *IEEE international joint conference on neural networks*. 2002.
9. Luckman, A., N.M. Allinson, A.M. Ellis, and B.M. Flude, "Familiar face recognition: A comparative study of a connectionist model and human performance.," *Neurocomputing*, 1995. **7**: p. 3-27.
10. Bellman, R.E., *Adaptive control processes: A guided tour*. 1961: Princeton University Press.
11. Jolliffe, I., T., *Principal component analysis*. 1986, New York: Springer-Verlag.
12. Kohonen, T., *Self organizing maps*. 3rd ed. 2001: Springer-Verlag.
13. Demartines, P. and J. Herault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets.," *IEEE Transactions on Neural Networks*, 1997. **8**no.(1): p. 148-154.
14. Cortes, C. and V. Vapnik, "Support-vector networks.," *Machine Learning*, 1995. **20**: p. 273-297.
15. Phillips, P.J., H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms.," *Image and Vision Computing*, 1998. **16**no.(5): p. 295-306.
16. Martiniz, A.M. and R. Benavente, "The AR face database.," Technical Report : 24, CVC. June, 1998.
17. Jesorsky, O., K. Kirchberg, and R. Frischholz. "Robust face detection using the hausdorff distance,". in *International Conference on Audio- and Video-based Biometric Person Authentication*. 2001. Halmstad, Sweden.
18. Grassberger, P. and I. Proccacia, "Measuring the strangeness of strange attractors.," *Physica D*, 1983. **9**: p. 189-208.