

Recognizing Facial Expressions: A comparison of Computational approaches

Aruna Shenoy, Tim M Gale^{1,2*}, Neil Davey^{1*},
Bruce Christiansen^{1*}, and Ray Frank^{1*}

^{1*} School of Computer Science, University of Hertfordshire,
United Kingdom, AL10 9AB

{a.i.shenoy, t.gale, r.j.frank, n.davey}@herts.ac.uk

^{2*}Department of Psychiatry, Queen Elizabeth II Hospital, Welwyn Garden City
Herts, AL7 4HQ UK

Abstract. Recognizing facial expressions are a key part of human social interaction, and processing of facial expression information is largely automatic, but it is a non-trivial task for a computational system. The purpose of this work is to develop computational models capable of differentiating between a range of human facial expressions. Raw face images are examples of high dimensional data, so here we use some dimensionality reduction techniques: Linear Discriminant Analysis, Principal Component Analysis and Curvilinear Component Analysis. We also preprocess the images with a bank of Gabor filters, so that important features in the face images are identified. Subsequently the faces are classified using a Support Vector Machine. We show that it is possible to differentiate faces with a neutral expression from those with a smiling expression with high accuracy. Moreover we can achieve this with data that has been massively reduced in size: in the best case the original images are reduced to just 11 dimensions.

Keywords: Facial Expressions, Image Analysis, Classification, Dimensionality Reduction.

1 Introduction

According to Ekman and Friesen [1] there are six easily discernible facial expressions: anger, happiness(smile), fear, surprise, disgust and sadness, apart from neutral. Moreover these are readily and consistently recognized across different cultures [2]. In the work reported here we show how a computational model can identify facial expressions from simple facial images. In particular we show how smiling faces and neutral faces can be differentiated. Data presentation plays an important role in any type of recognition. High dimensional data is normally reduced to a manageable low dimensional data set. We perform dimensionality reduction and classification using Linear Discriminant Analysis and also dimensionality reduction using Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection technique and it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) [3]. The Intrinsic Dimension (ID) [4], which is the true dimension of the data, is often much less than the original dimension of the data. To use this efficiently, the actual dimension of the data must be estimated. We use the Correlation Dimension to estimate

the Intrinsic Dimension. We compare the classification results of these methods with raw face images and of Gabor Pre-processed images [5],[6]. The features of the face (or any object for that matter) may be aligned at any angle. Using a suitable Gabor filter at the required orientation, certain features can be given high importance and other features less importance. Usually, a bank of such filters is used with different parameters and later the resultant image is a L2 max (at every pixel the maximum of feature vector obtained from the filter bank) superposition of the outputs from the filter bank.

2 Background

We begin with a simple experiment to classify two expressions: neutral and smiling. We use Linear Discriminant Analysis (LDA) for dimensionality reduction and classification. We also use a variety of other dimensionality reduction techniques, a Support Vector Machine (SVM) [7] based classification technique and these are described below.

2.1 Linear Discriminant Analysis (LDA)

For a two class problem, LDA is commonly known as Fisher Linear discriminant analysis after Fisher [8] who used it in his taxonomy based experiments. Belhuemer was the first to use the LDA on faces and used it for dimensionality reduction [9] and it can be used as a classifier. LDA attempts to find the linear projection of the data that produces maximum between class separation and minimum within class scatter. In the simple example shown in Figure 1, a projection on to the vertical axis separates the two classes whilst minimizing the within class scatter. Conversely, a projection onto horizontal axis does not separate the classes. Formally the algorithm can be described as follows. The between class scatter covariance matrix is given by:

$$\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \quad (1)$$

The within class covariance matrix is given by:

$$\mathbf{S}_W = \sum_{i=1}^C \sum_{n \in C_k} (\mathbf{X}^n - \mathbf{m}_i)(\mathbf{X}^n - \mathbf{m}_i)^T \quad (2)$$

where m_1 and m_2 are the means of the datasets of the class 1 and 2 respectively. C is the number of classes and C_k is the k_{th} class. The eigenvector solution of $\mathbf{S}_w^{-1}\mathbf{S}_B$ gives the projection vector which in the context of face image classification is known as the Fisher face.

2.2 Gabor Filters

A Gabor filter can be applied to images to extract features aligned at particular orientations. Gabor filters possess the optimal localization properties in both spatial and frequency domains, and they have been successfully used in many applications [10]. A

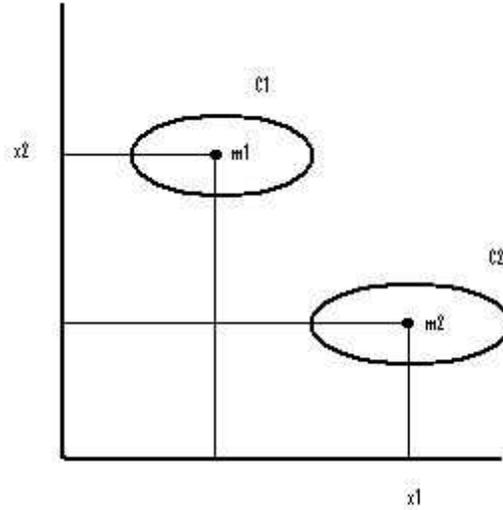


Fig. 1. The figure shows the classes which are overlapping along the direction of x_1 . However, they can be projected on to direction x_2 where there will be no overlap at all.

Gabor filter is a function obtained by modulating a sinusoidal with a Gaussian function. The useful parameters of a Gabor filter are orientation and frequency. The Gabor filter is thought to mimic the simple cells in the visual cortex. The various 2D receptive field profiles encountered in populations of simple cells in the visual cortex are well described by an optimal family of 2D filters [11]. In our case a Gabor filter bank is implemented on face images with 8 different orientations and 5 different frequencies. Recent studies on modeling of visual cortical cells [12] suggest a tuned band pass filter bank structure. Formally, the Gabor filter is a Gaussian (with variances S_x and S_y along x and y -axes respectively) modulated by a complex sinusoid (with centre frequencies U and V along x and y -axes respectively) and is described by the following equation 3

$$g(x, y) = \frac{\exp \left[-\frac{1}{2} \left[\left(\frac{x}{S_x} \right)^2 + \left(\frac{y}{S_y} \right)^2 \right] + 2\delta j(Ux + Vy) \right]}{2\delta S_x S_y} \quad (3)$$

The variance terms and dictates the spread of the band pass filter centered at the frequencies U and V in the frequency domain. This filter is complex in nature.

A Gabor filter can be described by the following parameters: The S_x and S_y of the Gaussian explain the shape of the base (circle or ellipse), frequency (f) of the sinusoid, orientation (θ) of the applied sinusoid Figure 2 shows examples of various Gabor filters. Figure 3b) shows the effect of applying a variety of Gabor filters shown in Figure 2 to the sample image shown in Figure 3a). Note how the features at particular orientations are exaggerated.

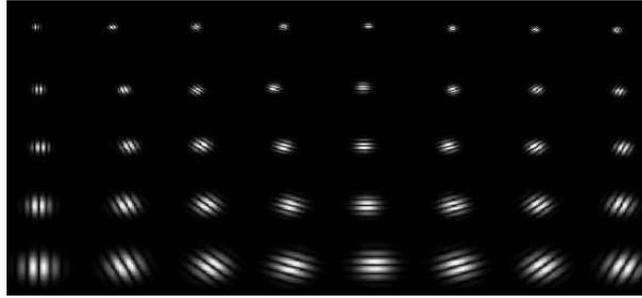


Fig. 2. Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

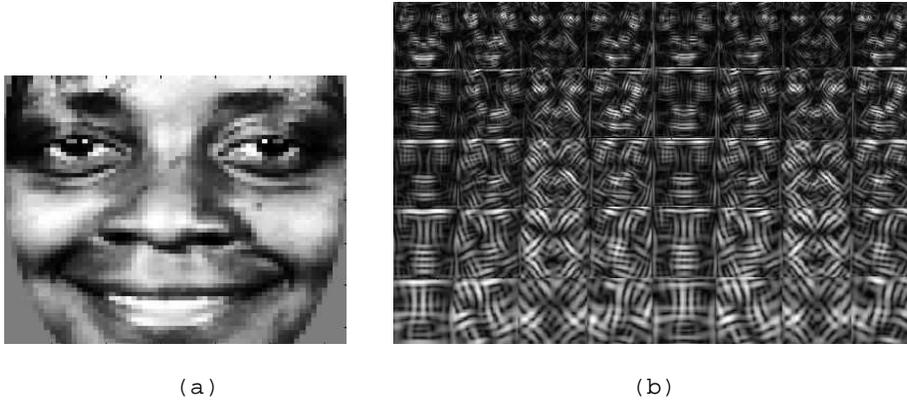


Fig. 3. (a) Original face image, (b) Forty Convolution outputs of Gabor

An augmented Gabor feature vector is created of a size far greater than the original data for the image. Every pixel is then represented by a vector of size 40 and demands dimensionality reduction before further processing. So a 64×64 image is transformed to size $64 \times 64 \times 5 \times 8$. Thus, the feature vector consists of all useful information extracted from different frequencies, orientations and from all locations, and hence is very useful for expression recognition.

Once the feature vector is obtained, it can be handled in various ways. We simply take the L2 max norm for each pixel in the feature vector. So that the final value of a pixel is the maximum value found by any of the filters for that pixel. The L2 max norm Superposition principle is used on the outputs of the filter bank and the Figure 4b) shows the output for the original image of Figure 4 a).

2.3 Curvilinear Component Analysis

Curvilinear Component Analysis (CCA) is a non-linear projection method that preserves distance relationships in both input and output spaces. CCA is a useful method

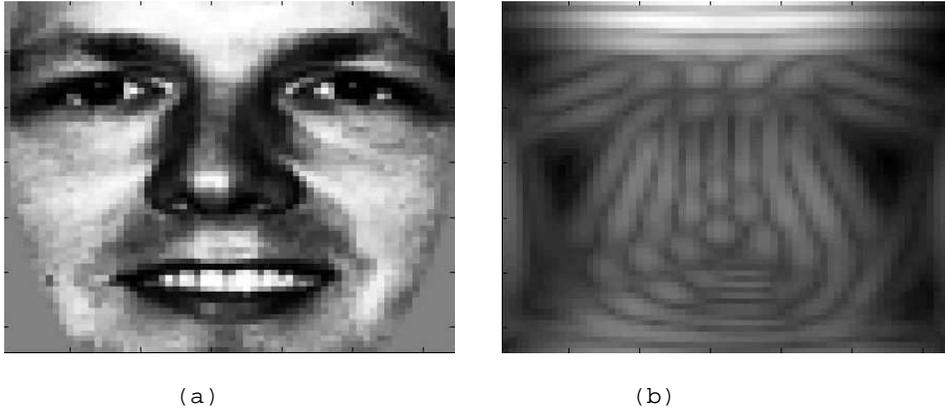


Fig. 4. a) Original Image used for the Filter bank b)Supposition Output(L2 max norm)

for redundant and non linear data structure representation and can be used in dimensionality reduction. CCA is useful with highly non-linear data, where PCA or any other linear method fails to give suitable information [3]. The D-dimensional input X should be mapped onto the output p-dimensional space Y . Their d-dimensional output vectors y_i should reflect the topology of the inputs x_i . In order to do that, Euclidean distances between the x_i 's are considered. Corresponding distances in the output space y_i 's is calculated such that the distance relationship between the data points is maintained.

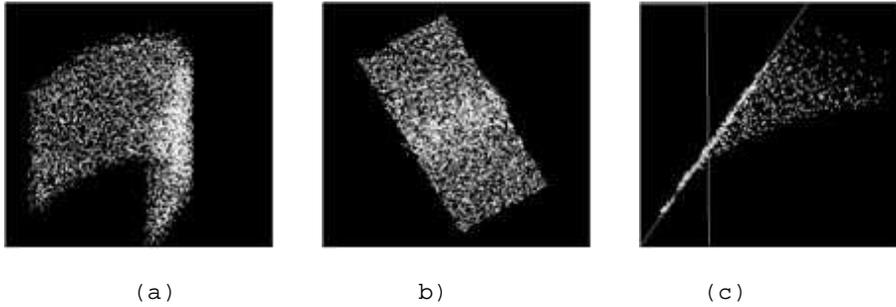


Fig. 5. (a) 3D horse shoe dataset (b) 2D CCA projection (c) plot.

CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N [d_{i,j}^X - d_{i,j}^Y]^2 F_{\lambda} [d_{i,j}^Y] \quad \forall j \neq i \quad (4)$$

where $d_{i,j}$ and $d_{i,j}^Y$ are the Euclidean distances between the points i and j in the input space X and the projected output space Y respectively and N is the number of data points. $F_\lambda [d_{i,j}^Y]$ is the neighbourhood function, a monotonically decreasing function of distance. In order to check that the relationship is maintained a plot of the distances in the input space and the output space ($d_y - d_x$) plot is produced. For a well maintained topology, d_y should be proportional to the value of d_x at least for small values of d_y 's. Figure 5 shows CCA projections for the 3D data horse shoe data. The ($d_y - d_x$) plot shown is good in the sense that the smaller distances are very well matched [5].

2.4 Intrinsic Dimension

One problem with CCA is deciding how many dimensions the projected space should occupy, and one way of obtaining this is to use the intrinsic dimension of the data manifold. The Intrinsic Dimension (ID) can be defined as the minimum number of free variables required to define data without any significant information loss. Due to the possibility of correlations among the data, both linear and nonlinear, a D -dimensional dataset may actually lie on a d -dimensional manifold ($D \geq d$). The ID of such data is then said to be d . There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

3 Classification using Support Vector Machines

A number of classifiers can be used in the final stage for classification. We have concentrated on the Support Vector Machine. Support Vector Machines (SVM) are a set of related supervised learning methods used for classification and regression. SVM's are used extensively for many classification tasks such as: handwritten digit recognition [14] or Object Recognition [15]. A SVM implicitly transforms the data into a higher dimensional data space (determined by the kernel) which allows the classification to be accomplished more easily. We have used the LIBSVM tool [7] for SVM classification.

4 Experiments and Results

We experimented on 120 faces (60 male and 60 female) each with two classes, namely: neutral and smiling (60 faces for each expression). The images are from The FERET dataset [16] and some examples are shown in Figure 76. The training set was 80 faces (with 40 female, 40 male and equal numbers of them with neutral and smiling). Two test sets were created. In both test sets the number of each type of face is balanced. For example, there were 5 smiling male faces and 5 smiling female faces. The first set has images which are easily discernible smiling faces. The second test set has smiling and neutral faces, but the smiling faces are not easily discernible. With all faces aligned based on their eye location, a 128 x 128 image was cropped from the original (150 x 130). The resolution of these faces is then reduced to 64 x 64.

A LDA projection was made onto the Fisher face shown in Figure 7. The two test sets were then classified by using the nearest neighbor in the test set in the projection



Fig. 6. Example FERET images used in our experiments which are cropped to the size of 128 x 128 to extract the facial region and reduced to 64 x 64 for all experiments.

Table 1. Classification accuracy of raw faces using LDA.

Accuracy %	Test Set 1	Test Set 2
LDA	95	75

space. The results are as in Table 1. Figure 7 shows the Fisher face obtained by performing the LDA on the training data set. 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. This resulted in using 66 components of the raw dataset and 35 components in the Gabor pre-processed dataset. As CCA is a highly non-linear dimensionality reduction technique, we use the intrinsic dimensionality technique and reduce the components to its Intrinsic Dimension. The Intrinsic Dimension of the raw faces was approximated as 14 and that of Gabor pre-processed images was 11. The classification results are shown in Table 2. Figure 8 shows the Eigenfaces obtained by the PCA technique.

After dimensionality reduction a standard SVM (with Gaussian kernel) was used to classify the images. The parameters of the SVM were optimized using 5-fold validation.

The results of the classification are as in Table 2. The PCA, being a linear dimensionality reduction technique, did not do quite as well as CCA. With CCA there was good generalization, but the key point to be noted here is the number of components used for the classification. The CCA makes use of just 14 components with raw faces and just 11 components with the Gabor pre-processed images to get good classification results.

This suggests that the Gabor filters are highlighting salient information which can be encoded in a small number of dimensions using CCA. Some examples of misclassifications are shown in Figure 9. The reason for these misclassifications is probably due to the relatively small size of training set. For example, the mustachioed face in the

Table 2. SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

SVM Accuracy %	Test Set 1	Test Set 2
Raw Faces(64x64)	95	80
Raw with PCA66	90	75
Raw with CCA14	90	80
Gabor pre-processed Faces (64x64)	95	80
Gabor with PCA35	70	60
Gabor with CCA11	95	80

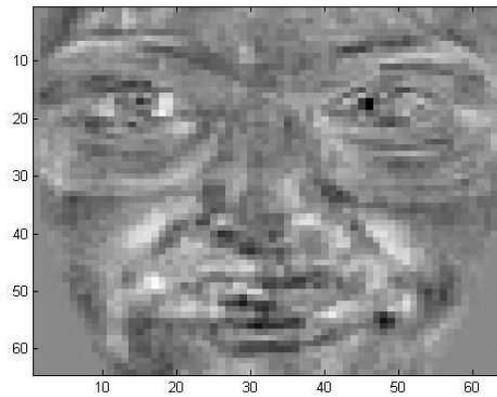


Fig. 7. Fisher face obtained for a dataset with two classes namely, Neutral and Smiling.



Fig. 8. The first 5 eigenfaces of the complete data set.

middle of the bottom row is misclassified as smiling. The only mustachioed face in the training set is of the same man smiling.

5 Conclusion

Identifying facial expressions is a challenging and interesting task. Our experiment shows that identification from raw images can be performed very well. However, with a larger data set, it may be computationally intractable to use the raw images. It is therefore important to reduce the dimensionality of the data. Performing classification using



Fig. 9. Examples of the misclassified set of faces. Top row shows smiling faces wrongly classified as neutral. Bottom row shows neutral faces wrongly classified as smiling.

LDA was a trivial task and the result was very impressive. It is interesting to see the effect size for each pixel in the image.

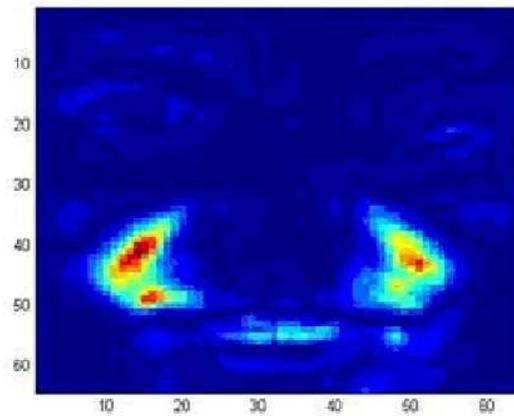


Fig. 10. Encoding face.

In other words which pixels discriminate most between smile and neutral faces can be seen and the result of this analysis is shown in Figure 10. The Creasing of the cheeks is diagnostic of smiling faces; teeth may also be an important indicator, though to a lesser extent. A linear method such as PCA does not appear to be sufficiently tunable to identify features that are relevant for facial expression characterization. Though the result of classification with LDA is impressive, for large datasets with face images, PCA

needs to be done prior to the LDA. However, on performing Gabor preprocessing on the images and following it with the CCA, there was good generalization in spite of the massive reduction in dimensionality. The most remarkable finding in this study is that the facial expression can be identified with just 11 components found by CCA. Future work will include extend the experiment to a larger data set and for other expressions.

References

1. Ekman, P. and W.V. Friesen, Constants across cultures in the face of the emotion. *Journal of Personality and Social Psychology*, 1971. 17.
2. Batty, B., M.J. Taylor, and, Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, 2003. 17.
3. Demartines, P. and D.J. Herault, Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 1997. 8(1): p. 148-154.
4. Grassberger, P. and I. Proccacia, Measuring the strangeness of strange attractors. *Physica D*, 1983. 9.
5. Jain, A.K. and F. Farrokhnia, Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 1991. 24(12).
6. Movellan, J.R., Tutorial on Gabor Filters. 2002.
7. Chang, C.-C. and Chih-Jen Lin, LIBSVM: a library for support vector machines. 2001.
8. Fisher, R.A., The use of mutiple measures in anatomical problems. *Ann. Eugenics*, 1936. 7: p. 179-188.
9. Belhumeur and Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on patternAnalysis and Machine Intelligence*, 1997. 19(7): p. 711-720.
10. Zheng, D., Y. Zhao, and J. Wang, Features Extraction using A Gabor Filter Family. *Proceedings of the sixth Lasted International conference, Signal and Image processing, Hawaii*, 2004.
11. Daugman, J.G., Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters. *Journal of Optical.Society of.America.A*, 1985. 2(7).
12. Kulikowski, Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics*, 1982. 43(3): p. 187-198.
13. Smith, L.I., Tutorial on Principal Component Analysis. 2002.
14. Cortes, C. and V. Vapnik, Support Vector Networks. *Machine Learning*, 1995. 20: p. 273-297.
15. Blanz, V., et al., Comparison of view-based object recognition algorithms using realistic 3D models. *Proc. Int. Conf. on Artificial Neural Networks*, 1996: p. 251-256.
16. Philips, P.J., et al., The FERET evaluation methodology for face recognition algorithms. *Image and Vision Computing*, 1998. 16(5): p. 295-306.