

# REPRESENTATION AND CLASSIFICATION OF FACIAL EXPRESSION IN A MODULAR COMPUTATIONAL MODEL

ARUNA SHENOY<sup>1</sup>  
TIM GALE<sup>1,2</sup>  
RAY FRANK<sup>1</sup>  
NEIL DAVEY<sup>1</sup>

*<sup>1</sup>Department of Computer Science, University of Hertfordshire,  
College Lane, Hatfield, AL10 9AB, UK*

*<sup>2</sup>Department of Psychiatry, QEII Hospital, Welwyn Garden City, AL7 4HQ, UK*

Recognizing expressions is a key part of human social interaction; Processing of facial expression information is largely automatic in humans, 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. Here we use two sets of images, namely: Angry and Neutral. Raw face images are examples of high dimensional data, so here we use some dimensionality reduction techniques: Principal Component Analysis and Curvilinear Component Analysis. We 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 also find the effect size of the pixels for the Angry and Neutral faces. We show that it is possible to differentiate faces with a neutral expression from those with an angry 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 6 dimensions.

## 1. Introduction

According to Ekman and Friesen [1] there are six easily discernible facial expressions: anger, happiness(smile), fear, surprise, disgust and sadness. 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. Specifically we investigate the differentiation of angry from neutral faces. In particular we show how angry 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 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 basically perform an experiment to classify two expressions: neutral and Angry. We do pre-processing by Gabor filters and dimensionality reduction by techniques, namely, Principal Component Analysis and Curvilinear Component Analysis followed by a Support Vector Machine (SVM) [7] based classification technique and these are described below.

### **2.1. 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 [8]. A 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 [9]. 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 [10] 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 Equation 1:-

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

(1)

The variance terms  $S_x$  and  $S_y$  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 ( $\Theta$ ) of the applied sinusoid. Figure 1 shows examples of various Gabor filters. Figure 2b) shows the effect of applying a variety of Gabor filters shown in Figure 1 to the sample image shown in Figure 2 a). Note how the features at particular orientations are exaggerated.

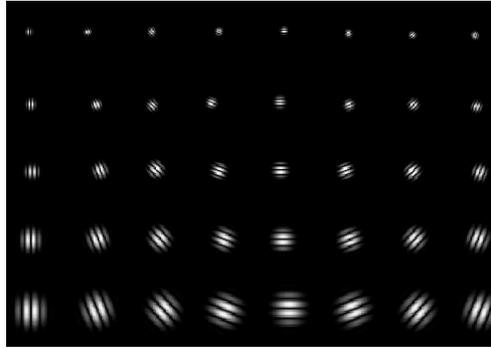


Figure 1: Gabor filters: Real part of the Gabor kernels at five scales and eight orientations

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  $63 \times 63$  image is transformed to size  $63 \times 63 \times 5 \times 8$ . 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.

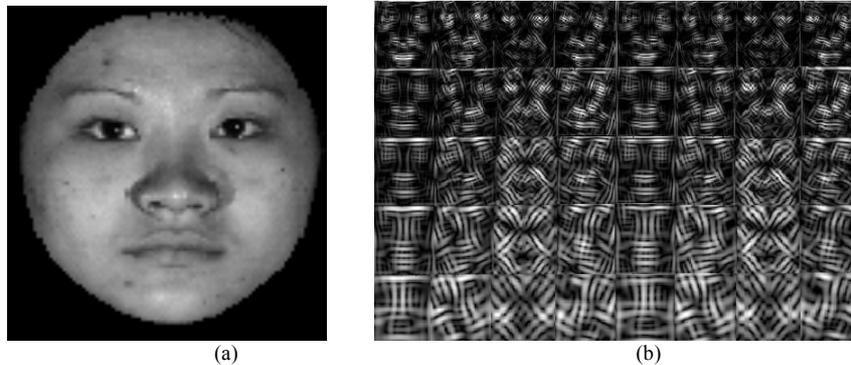


Figure 2: a) Original face image. b) Forty Convolution outputs of Gabor: The rows correspond to decreasing frequency (from top to bottom) and columns represent various orientation.

Once the feature vector is obtained, it can be handled in various

ways. We simply take the  $L_2$  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  $L_2$  max norm Superposition principle is used on the outputs of the filter bank and the Figure 3 b) shows the output for the original image of Figure 3 a).

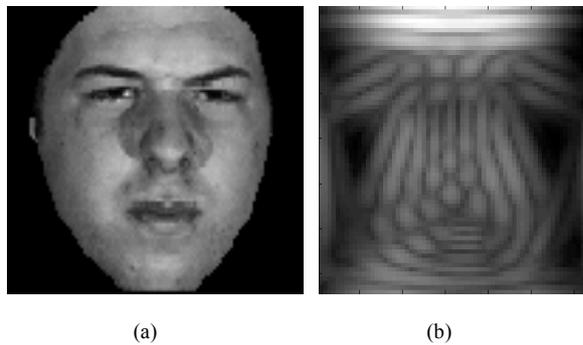


Figure 3 a): Original Image used for the Filter bank b) Superposition output ( $L_2$  max norm)

## 2.2. Principal Component Analysis

Principal Component Analysis (PCA) transforms higher dimensional datasets into lower dimensional uncorrelated outputs by capturing linear correlations among the data, and preserving as much information as possible in the data. PCA transforms data from the original coordinate system to the principal axes coordinate system such that the principal axis passes through the maximum possible variance in the data. The second principal axis passes through the next largest possible variance and this is orthogonal to the first axis. This is repeated for the next largest possible variances and so on. All these axes are orthogonal to each other. On performing this PCA on the high dimensional data, Eigenvalues or principal components are thus obtained [11]. The required dimensionality reduction is obtained by retaining only the first few principal components. Figure 4 shows the first two

principal components.

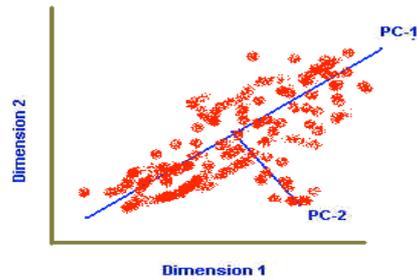


Figure 4: The first two consecutive principal components are shown.

### 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 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  $Y$ , where  $P \ll D$ . Their  $P$ - 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. CCA puts more emphasis on maintaining the short distances than the longer ones. Formally, this reasoning leads to the following error function:

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

where  $d_{i,j}^X$  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.  $F(d_{i,j}^Y)$  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 ( $dy - dx$  plot) is produced. For a well maintained topology,  $dy$  should be proportional to the value of  $dx$  at least for small values of  $dy$ 's. Figure 5 shows CCA projections for the 3D data horse shoe data. The  $dy - dx$  plot shown is good in the sense that the smaller distances are very well matched [3].

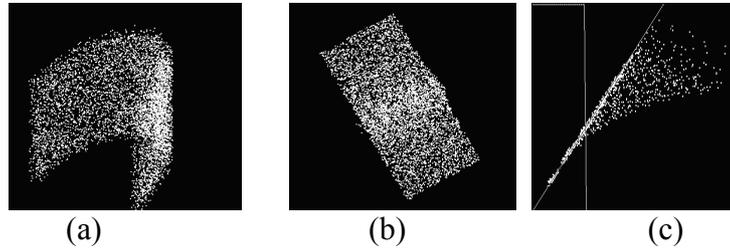


Figure 5: (a) 3D horse shoe dataset (b) 2D CCA projection (c)  $dy - dx$  plot.

#### 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  $P$ -dimensional manifold ( $D \geq P$ ). The ID of such data is then said to be  $P$ . There are various methods of calculating the ID; here we use the correlation Dimension [8] to calculate the ID of face image dataset.

### 2.5. *Encoding Face*

'Effect Size' is a way of expressing the difference between two groups. Here two groups: Angry and Neutral are used. Cohen [12] defined  $d$  as the difference between the means,  $M_1 - M_2$ , divided by standard deviation,  $\sigma$  of either group.

$$d = \frac{M_1 - M_2}{\sigma}$$

(3)

$M_1$  and  $M_2$  are the means of two groups and  $\sigma$  is the standard deviation and it is calculated by Equation 4.

$$\sigma = \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)}{N}} \quad (4)$$

$\sigma_1$  and  $\sigma_2$  are the standard deviation of the two classes, Angry and Neutral respectively and  $N$  is the total number of samples. 'Encoding face' is obtained by finding the Effect size of each pixel in an image. In other words it shows which pixels discriminate most between Angry and Neutral faces.

### 2.6. *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 Machine (SVM) is 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 [13] or Object Recognition [14]. 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.

### 3. Experiments and Results

We experimented on 200 faces (112 female and 88 male) each with two classes, namely: *Neutral* and *Angry* (100 faces for each expression). The images are from The BINGHAMTON dataset [15] and some examples are shown in Figure 6. The training set had 160 faces (with 46 female, 34 male and equal numbers of them with neutral and angry expression). The original  $128 \times 128$  image was reduced to  $63 \times 63$ . The test set consists of 40 faces (10 female, 10 male and equally balanced number of expression).



Figure 6: Examples of BINGHAMTON images used in our experiments was converted to gray scale and then reduced to size  $63 \times 63$  for all experiments.

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 105 components of the raw dataset and 22 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 10 and that of Gabor pre-processed images was 6. The SVM classification results are shown in Table 1. Figure 7 shows the Eigen faces obtained by performing the PCA on the data set. Figure 8 shows the  $dy-dx$  plot of the CCA projection for the data set.

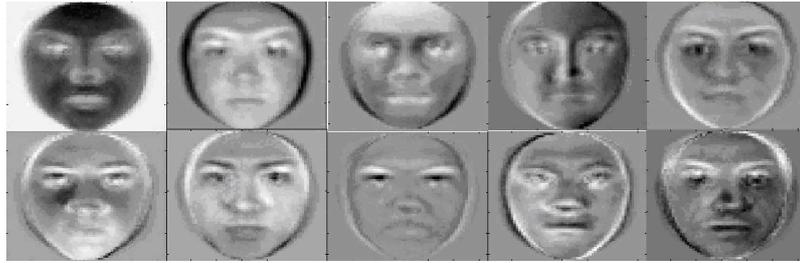


Figure 7: First ten Eigen faces of the dataset with two classes namely, Neutral and Angry.

The SVM was trained in the following way:

- 1) Transform the data to a format required for using the SVM software package - LIBSVM -2.86 [7].
- 2) Perform simple scaling on the data so that all the features or attributes are in the range  $[-1, +1]$ .
- 3) Choose a kernel. We used the RBF kernel,  

$$k(x, y) = e^{-\gamma \|x-y\|^2}.$$
- 4) Perform fivefold cross validation with the specified kernel to find the best values of the cost parameter  $C$  and  $\gamma$ .
- 5) Using the best value of  $C$  and  $\gamma$ , train the model and finally evaluate the trained classifier using the test sets.

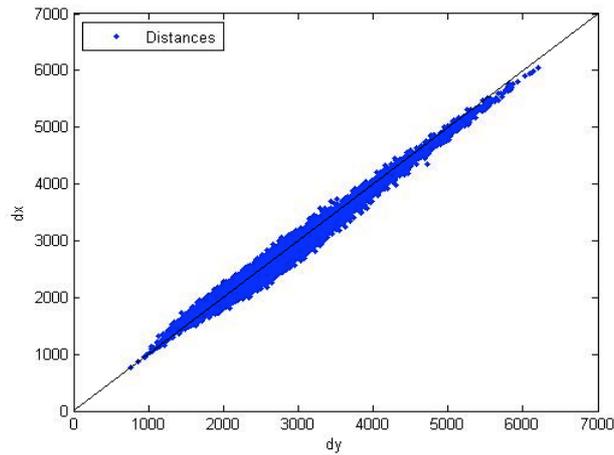


Figure 8: The  $dy$ - $dx$  plot of the CCA projection for the raw data set. If there is a good matching between input and output spaces and the data is linear, then all the distances would be on the line  $dy = dx$  line. Here the original 3969 dimensions have been reduced to just 10 components by CCA.

Table 1. SVM Classification accuracy of raw faces and Gabor pre-processed images with PCA and CCA dimensionality reduction techniques.

% SVM Accuracy	Testset (40 images)
Raw faces	37 (92.5%)
Raw with PCA105	27 (67.5%)
Raw with CCA10	31 (77.5%)
Gabor pre-processed faces	29 (72.5%)
Gabor with PCA22	30 (75%)
Gabor with CCA6	28(70%)

The Encoding Angry face, the image where pixels which discriminate most between Angry and Neutral faces, is shown in Figure 9. The eyebrows are pulled together and down to form

vertical wrinkles between the eyebrows in the forehead which is diagnostic of angry faces and can be seen clearly in the image. The glaring stare which is caused by the tightening of the muscles around the eyelids can also be somewhat seen [16]. The flaring of the nostrils and the clenching of jaws [17] may also be an important indicator, though to a lesser extent.

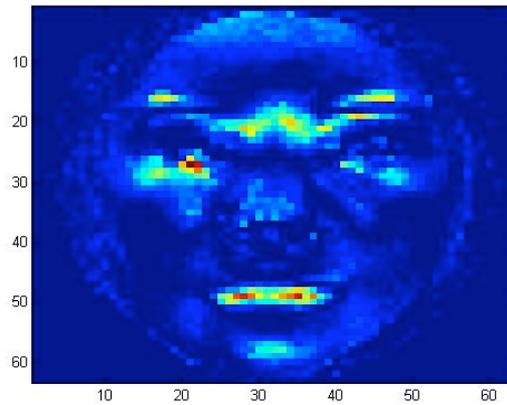


Figure 9: Encoding face: for Angry and Neutral

#### 4. Conclusions

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. The experiments so far have shown that Gabor pre-processed images, with dimensionality reduced by CCA to just 6 components, offer a promising approach for investigation. In order to examine the consistency of the different models, further experiments need to be run with larger datasets and with other expression categories. The Similarities and differences in these results may be useful and informative in developing a better computational model and may contribute to our understanding of human processing of face expressions. Also, the performance of the computational model will have to be compared with the performance accuracy of human subjects with respect to a range of 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. Héroult, *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. Procaccia, *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. 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.
9. 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).
10. 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.
11. Smith, L.I., *Tutorial on Principal Component Analysis*. 2002.
12. Cohen, J., ed. *Statistical power analysis for the behavioural sciences*. 1988, Lawrence Earlbaum Associates.: Hillsdale, New Jersey, .
13. Cortes, C. and V. Vapnik, *Support Vector Networks*. Machine Learning, 1995. **20**: p. 273-297.
14. 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.
15. Yin, L., Wei, X., Sun, Y., Wang, J. & Rosato, M. J. , *A 3D Facial Expression Database For Facial Behavior Research*. 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006
16. Hager, J.C. *Data Face*. 2003 [cited; Available from: <http://www.face-and-emotion.com/dataface/expression/interpretations.html>].
17. Novaco, R.W., *Anger*. Encyclopedia of Psychology. 2000: Oxford University Press.