A Comparison of the performance of humans and computational models in the classification of facial expression

A.Shenoy[1] (a.1.shenoy@herts.ac.uk)
Sue Anthony[2] (s.1.anthony@herts.ac.uk),
R.J.Frank[1] (r.j.frank@herts.ac.uk),
Neil Davey [1] (n.davey@herts.ac.uk)
[1] Department of Computer Science, [2] Department of Psychology,
University of Hertfordshire, College Lane, Hatfield
AL10 9AB

Keywords: Facial Expressions, Image Analysis, Classification, Reaction Time.

Abstract

Recognizing expressions are a key part of human social interaction, and processing of facial expression information is largely automatic for humans, but it is a non-trivial task for a computational system. In the first part of the experiment, we develop computational models capable of differentiating between two human facial expressions. We perform pre-processing by Gabor filters and dimensionality reduction using the methods: Principal Component Analysis, and Curvilinear Component Analysis. Subsequently the faces are classified using a Support Vector Machines. We also asked human subjects to classify these images and then we compared the performance of the humans and the computational models. The main result is that for the Gabor pre-processed model, the probability that an individual face was classified in the given class by the computational model is inversely proportional to the reaction time for the human subjects.

Introduction

In this work we compare the performance of human subjects classifying facial expressions, with the performance of a variety of computational models. We use a set of 176 face images, half of which express anger and the other half have a neutral expression. The images are from the BINGHAMTON BU-3DFE database (Yin, Wei et al. 2006) and some examples are shown in Figure 1.

Pre-Processing Methods and Classification

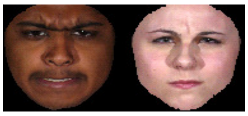
This section describes how the computational model classifies angry faces and neutral faces. High dimensional data such as face images are often reduced to a more manageable low dimensional data set. We perform dimensionality reduction using both Principal Component Analysis (PCA) and Curvilinear Component Analysis (CCA). PCA is a linear projection technique but it may be more appropriate to use a non linear Curvilinear Component Analysis (CCA) (Demartines and Hérault 1997). Gabor filters are also often used for extracting features of images, and they are thought to mimic some aspects of human visual processing (Daugman 1985). Classification is performed using a Support Vector Machines (SVM). An SVM performs classification by finding the maximum margin hyper-plane in a feature space. The relative distance of an instance from this hyper-plane can be interpreted as its probability of belonging to the appropriate class. We have used the LIBSVM-2.86 tool (Chang and Lin 2001).

Experiment

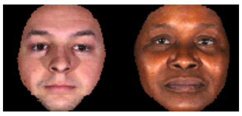
Two sets of experiments were performed. Part A - Computational models. Part B - Classification performed by human subjects.

Part A- Computational Models

The data was divided into four subsets, and training/testing took place with a leave one out strategy, so that results are averages over four independent runs. Once a training set had been selected the two parameters of the SVM were optimized by cross-validation. Six variations of data processing are tested as detailed in Table 1.



(a)



(b)

Figure 1: Example face images. a) Angry b) Neutral

Computational Model Results

For PCA, the first 97 components of the raw dataset and 22 components in the Gabor pre-processed dataset account for 95% of the total variance. For CCA, we reduce the data to its Intrinsic Dimension. The intrinsic dimension of the raw faces was approximated as 5 and that of the Gabor pre-processed images was 6.

The results in Table 2 indicate the overall classification accuracy is not very good; however, classifying angry faces is a difficult task for computation models (Susskind 2007) and can be seen from the results. Nevertheless, the SVM performs well with an average of 84.09% accuracy with raw face images

.

Table 1: Types of Computational Models

Name model	Type of Input	Dimensionality Reduction
Model 1	Raw faces	None
Model 2	Raw faces	PCA
Model 3	Raw faces	CCA
Model 4	Gabor pre-processed	None
Model 5	Gabor pre-processed	PCA
Model 6	Gabor pre-processed	CCA

Table 2: SVM classification Results

Accuracy	TEST SET 4	TEST SET 3	TEST SET 2	TEST SET 1	Average
Model 1	79.54% (35/44)	93.18% (41/44)	79.54% (35/44)	84.09% (37/44)	84.09%
Model 2 (PCA97)	68.18% (30/44)	77.27% (34/44)	70.45% (31/44)	65.91% (29/44)	70.45%
Model 3 (CCA5)	68.18% (30/44)	59.09% (26/44)	63.64% (28/44)	63.64% (28/44)	63.64%
Model 4	68.18% (30/44)	79.55% (35/44)	72.73% (32/44)	81.82% (36/44)	75.57%
Model 5 (PCA22)	61.36% (27/44)	79.55% (35/44)	75% (33/44)	72.73% (32/44)	72.16%
Model 6 (CCA6)	63.64% (28/44)	70.45% (31/44)	68.18% (30/44)	63.64% (28/44)	66.48%

Part B - Human subjects

The 184 raw images were used in this experiment. Twenty individuals took part in the study.

Method

A total of 16 images were used in the pre-view block and the remaining 168 images were divided into 6 balanced blocks of 28 images each. We used a tool called as TESTBED (Taylor 2003) which is a response test generator program to record the classification and the Response Time (RT) of individuals.

Human Subject Results

Humans correctly classified the target expression with a mean of 82.86% (SD = 0.174) and the average RT was 1.132 seconds (SD = 0.714). The average RT ranges between a maximum value of 1.792sec and a minimum value of 0.714sec.

Discussion

We use the Bi-Variate Correlation to find any correlation between the average RT for human subjects and the class membership probability for the computational models. The results are considered to be significant at the level of 0.05, or below. The results of comparison are shown in correlation matrix of Table 3.

Table 3: The Bi-Variate Correlation Results

Model	Correlation value	Significance value
Model 1	-0.005	0.391
Model 2	+0.002	0.645
Model 3	-0.022	0.126
Model 4	-0.045	0.016
Model 5	-0.028	0.065
Model 6	-0.003	0.597

Interestingly all but one of the correlations are negative, but only for Model 4 (Gabor filtered images with no dimensionality reduction) was this correlation significant, with the probability of the null hypothesis being 0.016. The correlation is negative with value -0.045. This negative correlation indicates large average RT (which presumably indicates that the subjects found it hard to classify), correlates with smaller class membership probability for the model. The results are interesting and encouraging (suggestive of Gabor filtering is similar to human face processing) and our next step is extending these experiments to other expressions.

References

Chang, C.-C. and C.-J. Lin (2001). "LIBSVM: a library for support vector machines."

Daugman, J. G. (1985). "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two dimensional visual cortical filters." *Journal of Optical.Society of.America .A* **2**(7).

Demartines, P. and D. J. Hérault (1997). "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets " *IEEE Transactions on Neural Networks* **8**(1): 148-154.

Susskind, J. M., G. Littlewort, M.S. Bartlett, J. Movellan,A.K. Anderson((2007). "Human and computer recognition of facial expressions of emotion." *Neuropsychologia* **45**(1).

Taylor, N. (2003). Developing with Authorware- Test bed. *ATSiP Conference at the University of Hertfordshire*

Yin, L., X. Wei, et al. (2006). A 3D Facial Expression Database For Facial Behavior Research. *7th International Conference on Automatic Face and Gesture Recognition* (FGR06).