JMLR: Workshop and Conference Proceedings 11 (2010) 111–118 Workshop on Applications of Pattern Analysis

Facial Expression Detection using Filtered Local Binary Pattern Features with ECOC Classifiers and Platt Scaling

Raymond S. Smith Raymond.Smith@surrey.ac.uk
Terry Windeatt T.Windeatt@surrey.ac.uk
Centre for Vision, Speech and Signal Processing,
University of Surrey, Guildford, Surrey GU2 7XH, UK

Editors: Tom Diethe, Nello Cristianini, John Shawe-Taylor

Abstract

We outline a design for a FACS-based facial expression recognition system and describe in more detail the implementation of two of its main components. Firstly we look at how features that are useful from a pattern analysis point of view can be extracted from a raw input image. We show that good results can be obtained by using the method of local binary patterns (LPB) to generate a large number of candidate features and then selecting from them using fast correlation-based filtering (FCBF). Secondly we show how Platt scaling can be used to improve the performance of an error-correcting output code (ECOC) classifier.

Keywords: Face expression recognition, local binary patterns, feature selection, multi-classifier systems

1. Introduction

Automatic face expression recognition is an increasingly important field of study that has applications in several areas such as human-computer interaction, human emotion analysis, biometric authentication and fatigue detection. One approach to this problem is to attempt to distinguish between a small set of prototypical emotions such as fear, happiness, surprise etc. In practice, however, such expressions rarely occur in a pure form and human emotions are more often communicated by changes in one or more discrete facial features. For this reason the facial-action coding system (FACS) Ekman and Friesen (1978); Tian et al. (2001) is commonly employed; in this method, expressions are characterised as groups of elementary facial movements known as action units (AUs). Some examples of AUs from the region around the eyes are shown in Fig. 1.

In this paper we discuss the design of FACS-based face expression recognition systems and examine in more detail some of the problems encountered. Section 2 presents an overview of the architecture of such a system and then sections 3 to 5 describe the results of experimental work that has been undertaken to investigate the merits of some possible implementation choices. For these experiments the Cohn-Kanade face expression database Kanade et al. (2000) was used. Finally, section 6 summarises the conclusions to be drawn from this work.

2. System Architecture

The overall architecture for an expression recognition system is illustrated in Fig. 2 and consists of the following elements:

Image capture consists of a camera system to capture face images. Depending on the application, this can be a video camera which allows for a real-time response to a continual stream of images or a still camera that obtains face images in a more controlled environment. A recent trend is also to capture three-dimensional images.

Face detection and registration is responsible for locating a face image within the input image and for determining the position of facial landmarks such as eye centres and the tip of the nose. This is a more or less difficult problem depending on how controlled are the conditions under which the image is captured. In very uncontrolled conditions, pose normalisation may also have to be performed.

Feature extraction and selection obtains useful attributes from the matrix of pixel intensities[1] which represents the raw input image. It is possible to directly detect AUs by measuring the relative position of a large number of facial landmarks Tian et al. (2001). It has been found, however, that comparable or better results can be obtained by taking a more holistic approach to feature extraction Donato et al. (1999) using methods such as Gabor wavelets, principal components analysis (PCA) Turk and Pentland (1991) and local binary patterns (LBP) Ahonen et al. (2006). These methods often produce a very large number of candidate features so some method must be used to select the more relevant ones (or synthesise new ones) from this set. PCA and LBP are discussed further in section 4.

AU classification applies some form of pattern analysis technology to make classification decisions about the presence or absence of individual AUs in the input image. Such a classifier must first be trained using a set of images for which the AUs have been manually determined. AU classification is discussed further in section 5.

Expression classification maps the set of detected AUs to one of the expressions which is of interest in the particular application. For common human emotions, such as fear, anger, surprise etc., this mapping can be performed using a standard FACS code book. In some applications, the definition of what constitute significant AU groupings will need to be determined on an application-specific basis.

Appropriate action means the action taken by an external system in response to the detected facial expression. Some examples are the sounding of an alarm if the driver of a vehicle is showing signs of drowsiness, offering help information to the user of a computer system who displaying an expression of puzzlement and raising an alert when a stressed or nervous person is attempting to enter a secure building.

3. Experimental Approach,

In the following two sections we present the results of performing classification experiments on the Cohn-Kanade face expression database. This dataset contains frontal video clips of posed expression sequences from 97 university students. Each sequence goes from neutral to target display but only the last image is AU coded.

In these experiments we focused on detecting AUs from the the upper face region as shown in Fig. 1. To detect individual AUs we adopted a two stage approach. Firstly a multi-class classifier was trained to recognise a number of commonly occurring AU groups

1. Colour images are usually represented by three such matrices, one each for the red, green and blue components. Greater accuracy can be achieved by combining these sources of information, but this is outside the scope of this paper, which focuses on greyscale images only.

Figure 1: Some example AUs and AU groups from the region around the eyes. AU1 = inner brow raised, AU2 = outer brow raised, AU4 = brows lowered and drawn together, AU5 = upper eyelids raised, AU6 = cheeks raised, AU7 = lower eyelids raised. The images are shown after manual eye location, cropping, scaling and histogram equalisation.

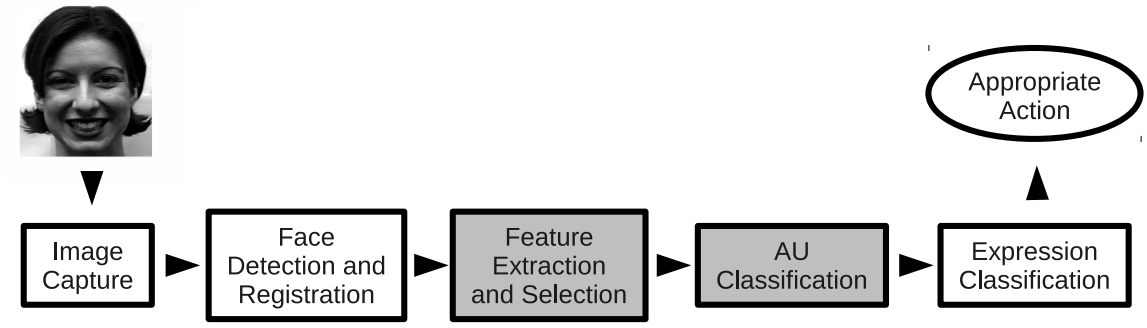

Figure 2: Face expression recognition system architecture. The two shaded elements are described in more detail in this paper.

as shown in Table 1. Different combinations of the soft outputs from this classifier were then used to obtain a score for each AU separately. For example, to detect AU2 the soft outputs for groups 2,3 and 12 were combined and compared with those for groups 1 and 4 to 11. Different combination methods were examined; these are described in section 5.

Neutral images were not used in these experiments and AU groups with three or fewer examples were ignored. In total this led to 456 images being available for training and testing. Note that researchers often make different decisions in these areas, and in some cases are not explicit about which choice has been made. This can render it difficult to make a fair comparison with previous results. For example some studies use only the last

Table 1: Action unit groups used in the experiments.

Group number	1	2	3	4	5	6	7	8	9	10	11	12
AUs present	None	1,2	1,2,5	4	6	1,4	1,4,7	4,7	4,6,7	6,7	1	1,2,4
Number of examples	152	23	62	26	66	20	11	48	22	13	7	6

image in the sequence but others use the neutral image to increase the numbers of negative examples. Furthermore, some researchers consider only images with single AU, whilst others use combinations of AUs. We consider the more difficult problem, in which neutral images are excluded and images contain combinations of AUs. A further issue is that some papers only report overall error rate. This may be misleading since class distributions are unequal, and it is possible to get an apparently low error rate by a simplistic classifier that classifies all images as non-AU. For this reason we also report the area under ROC curve for each AU, similar to Bartlett et al. (2006).

Processing of input images was as follows. For each 640 x 480 pixel image we converted to greyscale by averaging the RGB components and located the eye centres manually. A rectangular window around the eyes was obtained and then rotated and scaled to 150 x 75 pixels. Histogram equalization was used to standardise the image intensities.

Multi-class AU group classification was performed using the error-correcting output code (ECOC) technique Dietterich and Bakiri (1995). ECOC proved to be a highly successful way of solving a multiclass learning problem by decomposing it into a series of 2-class problems, or dichotomies, and training a separate base classifier to solve each one. These 2-class problems are constructed by repeatedly partitioning the set of target classes into pairs of super-classes so that, given a large enough number of such partitions, each target class can be uniquely represented as the intersection of the super-classes to which it belongs. The classification of a previously unseen pattern is then performed by applying each of the base classifiers so as to make decisions about the super-class membership of the pattern. Redundancy can be introduced into the scheme by using more than the minimum number of base classifiers and this allows errors made by some of the classifiers to be corrected by the ensemble as a whole.

In these experiments ECOC ensembles of size 200 were constructed with single hidden-layer MLP base classifiers trained using the Levenberg-Marquardt algorithm. A range of MLP node numbers (from 2 to 16) and training epochs (from 2 to 1024) was tried; each such combination was repeated 10 times and the results averaged. Each run was based on a different randomly chosen stratified training set with a 90/10 training/test set split. To increase diversity each base classifier was trained on a separate bootstrap replicate drawn from the common training set by repeated sampling with replacement to produce a set of the same size. Another source of random variation was the initial MLP network weights. A further enhancement to ECOC was to apply class separability weighting Smith and Windeatt (2010) when decoding the outputs from the base classifiers. The ECOC partitions were randomly chosen but in such a way as to have balanced numbers of AU groups in each one.

The experiments were programmed in Matlab and made use of the PRTools software Duin et al. (2007). They were carried out on a (non-dedicated) Dell host with 16 x 3 GHz Intel Xeon processors and 32 Gb of RAM. The typical time taken to perform a single train and test run was 25 minutes using the LBP method and 105 minutes for each fixed number of PCA features.

4. Feature Extraction Methods

As noted in section 2, one of the stages in processing an input face image is to extract features from the raw pixel array that are more suitable for the application of pattern analysis tools. In this section we examine two possibilities for this. The first of these is PCA in which the the image is projected onto a basis of 'eigenfaces'. The second method is to use LBP features. The latter is a computationally efficient texture description method that has the benefit that it is relatively insensitive to lighting variations. LBP has been successfully applied to facial expression analysis Shan et al. (2009); in our experiments

AU	Error(%)		Area Under ROC (%)	
	PCA	LBP+FCBF	PCA	LBP+FCBF
1	10.2	**8.7**	94.1	**94.6**
2	**5.0**	**5.0**	96.5	**96.8**
4	9.6	**8.7**	92.3	**96.2**
5	4.6	**3.9**	97.9	**98.1**
6	**11.2**	**11.2**	89.8	**93.0**
7	11.4	**9.2**	**93.2**	92.8
mean	8.7	**7.8**	94.0	**95.3**

Table 2: A comparison of two different image feature extraction methods for AU the recognition problem.

LBP features were extracted by computing a uniform (i.e. 59-bin) histogram for each subwindow in a non-overlapping tiling of the upper face image. This was repeated with a range of tile sizes (from 12 x 12 to 150 x 75 pixels) and sampling radii (from 1 to 10 pixels). The histogram bins were then concatenated to give 107,000 initial features

In order to reduce the number of features, a natural choice for PCA is to use only those features that account for most of the variance in the set of training images. For the LBP representation we adopt the very efficient fast correlation-based filtering (FCBF) Yu and Liu (2003) algorithm to perform this function. FCBF operates by repeatedly choosing the feature that is most correlated with class, excluding those features already chosen or rejected, and rejecting any features that are more correlated with it than with the class. As a measure of correlation, the information-theoretic concept of symmetric uncertainty is used. When applied to the LBP features, FCBF reduced their number from 107,000 down to 120.

Table 2 compares these two approaches in terms of AU classification error and area under the ROC curve. The evidence from this table is that LBP yields better results than PCA and is to be preferred. A further advantage of the LBP-based method is that the optimal number of features is automatically selected by FCBF, whereas for PCA this is a free parameter whose value must be separately determined by a method such as cross-validation.

5. AU Group Combination Methods

One problem to be solved when using the approach described in this paper is how to best combine the outputs from the AU groups classifier in order to make decisions about the presence or otherwise of individual AUs. This section compares the results from several possible methods of combination, which are as follows:

'sum': the simple sum of the raw classifier outputs for groups which contain the target AU is compared with the sum for groups which do not contain it.

'hard': the classifier outputs are first rounded to 0 or 1 before summing.

'wsum': the summed outputs are weighted by the inverse of the number of groups represented in the sum.

'csum': as for 'sum' but the resulting AU scores are then mapped to estimates of probability using the Platt scaling algorithm. Platt scaling Platt (1999) refers to a technique whereby a score-to-probability calibration curve is calculated using the training set.

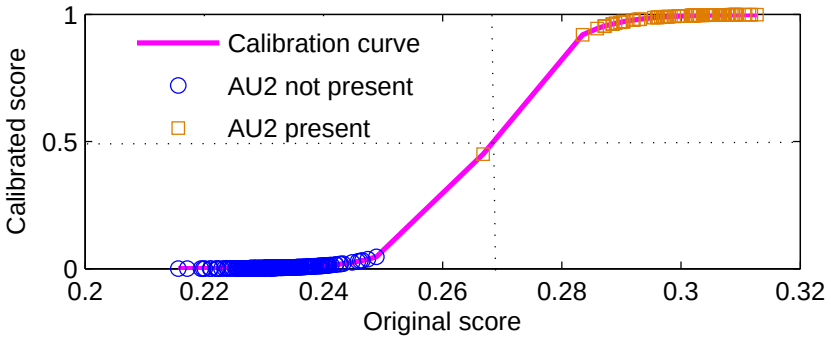

Figure 3: Calibration curve for AU2 training set (bootstrapping plus CSEP weighting applied).

This curve is based on the regularisation assumption that it is sigmoidal with equation $p(s) = \frac{1}{1+\exp(As+B)}$, where $p(s)$ is the probability that an image with score s contains the target AU. The parameters A and B together determine the slope of the curve and its lateral displacement; their values are determined by applying an expectation maximisation algorithm. A separate calibration curve is computed for each target AU.

An example of the kind of calibration curves that result from the Platt scaling algorithm is shown in Fig. 3 and the effect of applying the mapping to the test set is shown in Fig. 4. Note that, before calibration all scores are below 0.5 and hence would be classed as AU not present. After calibration (Fig. 4(b)) most of the test patterns that contain AU2 fall to the right hand side of the 0.5 threshold and hence are correctly classified.

The results of applying these different combination algorithms is shown in Fig.5 and it can be seen that the 'sum', 'wsum' and 'csum' methods lead to identical values for area under ROC curve. The reason for this is that the application of any monotonically increasing function to a score does not affect the shape of the ROC curve, it only affects the threshold values associated with each point on the ROC curve. The 'hard' algorithm, however makes discrete decisions and this alters the shape of the ROC curve in an adverse way.

As far as AU recognition error rates are concerned, the lowest values are obtained from 'csum', closely followed by 'hard'. The ideal to be aimed for is a high area under ROC curve and a low classification error rate. In this respect, 'csum' is the clear winner.

6. Discussion and Conclusions

In this paper we have outlined the design of a FACS-based system for detecting and responding to facial expressions and have described some of the problems to be solved in implementing such a system.

For image feature extraction, the LBP method leads to a very large number of candidate features but these can be reduced to a much smaller set of useful features by FCBF filtering. Evidence has been presented to show that this approach leads to better results than PCA.

The method adopted for AU detection was to train an ECOC classifier to simultaneously recognise commonly occurring AU groups and then to combine the soft outputs for each group to obtain an estimate of the probability of occurrence of each individual AU. Several methods for combining the soft outputs have been examined and it has been shown that

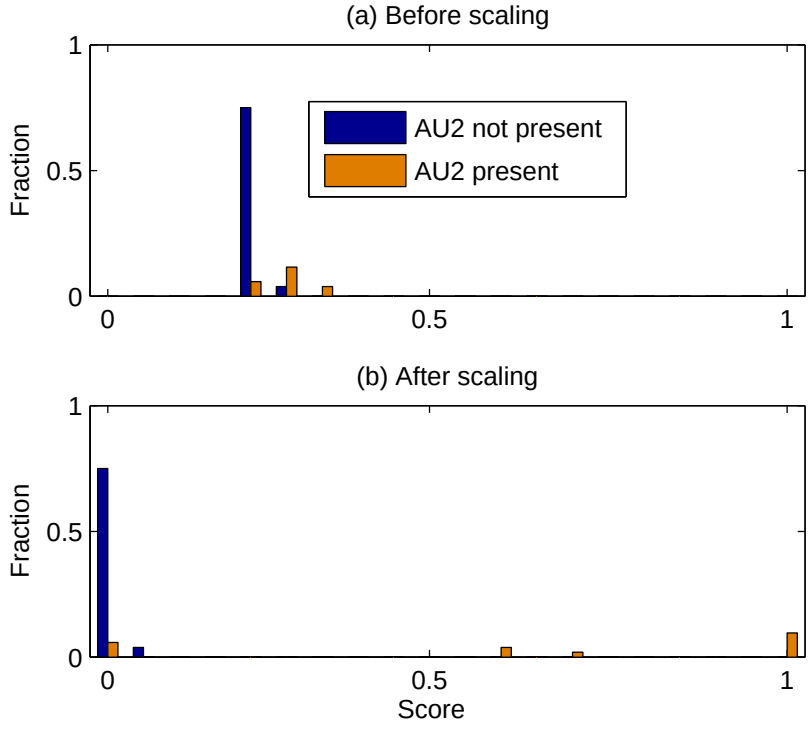

Figure 4: The effect of Platt scaling on the distribution of test-set scores for AU2.

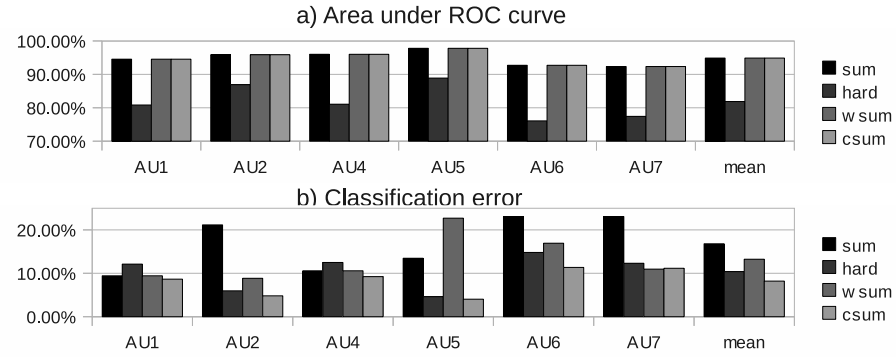

Figure 5: A comparison of different classifier combination algorithms.

the use of Platt scaling to first convert the resulting scores to probability estimates is an effective approach.

Experiments conducted on the Cohn-Kanade face expression database to find upper-face AUs have shown that good results can be achieved by these methods.

Another practical issue that must be borne in mind is that of efficiency. From this point of view, it is worth noting that both LBP and FCBF (which is only required during training) are fast lightweight techniques. The use of a single classifier, rather than one per AU, also helps to minimise the computational overheads of AU detection.

7. Acknowledgements

This work was supported by EPSRC grant E061664/1.

References

T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *PAMI*, 28:2037–2041, December 2006.

M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behaviour. In *Proc 7th Conf. On Automatic Face and Gesture Recognition*, pages 223–238, 2006.

T.G. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.

G. Donato, M.S. Bartlett, J.C. Hager, and T.J. Sejnowski P. Ekman. Classifying facial actions. *PAMI*, 21:974–989, October 1999.

R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, and S.Verzakov. *A Matlab Toolbox for Pattern Recognition*. Delft University of Technology, 2007.

P. Ekman and W.V. Friesen. *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*. San Francisco: Consulting Psychologists Press, 1978.

T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proc. 4th Int. Conf. Automatic Face and Gesture Recognition*, pages 46–53, March 2000.

J. Platt. *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA, 1999.

C. Shan, S. Gong, and P.W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27:803–816, June 2009.

R.S. Smith and T. Windeatt. Class-separability weighting and bootstrapping in error correcting output code ensembles. In *Proc. 9th Int. Conf. on Multiple Classifier Systems)*, volume 5997 of *LNCS*, pages 185–194, Berlin, 2010. Springer.

Y-I Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. *PAMI*, 23:97–115, February 2001.

M. Turk and A. Pentland. Eigenfaces for face recognition. *J. Cognitive Neuroscience*, 3:71–86, JAN 1991.

L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proc 12th Int Conf on Machine Learning (ICML-03)*, pages 856–863, 2003.