Neurocomputing 150 (2015) 440–448

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Facial action unit recognition using multi-class classification

R.S. Smith, T. Windeatt *

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7XH, UK

ARTICLE INFO

Article history:
Received 14 December 2013
Received in revised form
6 May 2014
Accepted 3 July 2014
Available online 5 October 2014

Keywords:
Facial expression recognition
Local binary patterns
Feature selection
Error-correcting output codes

ABSTRACT

Within the context of facial expression classification using the facial action coding system (FACS), we address the problem of detecting facial action units (AUs). Feature extraction is performed by generating a large number of multi-resolution local binary pattern (MLBP) features and then selecting from these using fast correlation-based filtering (FCBF). The need for a classifier per AU is avoided by training a single error-correcting output code (ECOC) multi-class classifier to generate occurrence scores for each of several AU groups. A novel weighted decoding scheme is proposed with the weights computed using first order Walsh coefficients. Platt scaling is used to calibrate the ECOC scores to probabilities and appropriate sums are taken to obtain separate probability estimates for each AU individually. The bias and variance properties of the classifier are measured and we show that both these sources of error can be reduced by enhancing ECOC through bootstrapping and weighted decoding.

1. Introduction

Automatic facial 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 driver fatigue detection [1,2]. One approach to this problem is to attempt to distinguish between a small set of prototypical emotions such as fear, happiness, and surprise based on the overall appearance of a face image [3–5]. 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) [6,7] is commonly employed to characterise individual facial movements as one of the 44 types known as action units (AUs). Facial expressions of interest (e.g. emotions or fatigue) can then be detected by looking for particular configurations of AUs; in this way the *interpretation* of facial expressions can be de-coupled from their detection. Some example of AUs from the region around the eyes are shown in Fig. 1. In this paper we refer to the first approach to expression recognition as *emotion recognition* and the second as *AU recognition*.

A commonly used approach to recognising facial AUs is to train a separate classifier to detect the presence or absence of each AU individually, for example [8]. One potential problem with this approach, however, is that AUs are not necessarily independent, as the presence of one AU may affect the appearance of another. They may also occur at different intensities and may occur on only one side of the face. A second possibility, which avoids these difficulties, is to take a holistic view of the problem by detecting commonly occurring AU groups directly. From a classification point of view the first approach treats the problem as a set of independent two-class problems while the second treats it as a single multi-class problem. Note that the second approach does not preclude the detection of individual AUs as, given a suitable classifier that outputs AU group probabilities, the probability of occurrence of a single AU can be obtained by summing the probabilities over all AU groups that contain it.

In this paper we adopt the second approach to AU recognition from still images. An important aspect of any image classification problem such as this, is the method used to extract relevant features from the image, and here we propose the method of multi-scale local binary patterns (MLBP) [9] to generate a large number of candidate features. To avoid problems resulting from the curse of dimensionality [10] this set must be reduced to a more manageable number, and for this purpose we employ the fast correlation-based filtering (FCBF) algorithm described in [11]. To the best of our knowledge this approach to feature extraction has not been previously applied to AU detection. In order to perform multi-class classification we make use of an error-correcting output code (ECOC) ensemble [12,13] of multi-layer perceptron (MLP) neural networks. Improved ECOC classification accuracy is achieved through the use of bootstrapping and a novel weighted decoding scheme. When bootstrap sampling of the training set is applied this is referred to as bootstrap aggregation or *bagging* and it has been shown to give improved results over non-bagged ensembles [14]. Furthermore it has been shown that this improvement in performance is brought about by a reduction in variance error [15,16]. Finally, as noted above, it is desirable that the scores

* Corresponding author.
E-mail address: T.Windeatt@surrey.ac.uk (T. Windeatt).

http://dx.doi.org/10.1016/j.neucom.2014.07.066
0925-2312/

Fig. 1. Some example of 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.

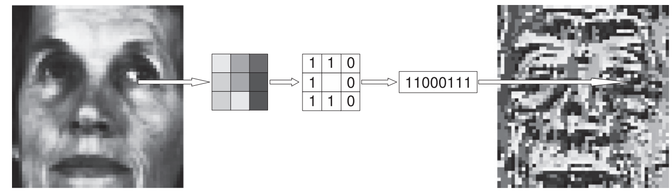

Fig. 2. Local binary pattern image production. Each non-border pixel is mapped as shown.

output by the classifier are an approximate measure of the probability of each AU group being present in the image. The standard ECOC classifier does not have this property and, to correct for this, we propose the use of Platt scaling [17] to post-process the scores and convert them to probability measures.

The method of computing the weights for weighted decoding ECOC is novel and based on spectral coefficients. The use of Walsh spectral coefficients for ensemble design was first reported in [18] for two-class problems. In [19] the relationship between second order Walsh coefficients and the Tumer–Ghosh model was established, and later applied to ensemble pruning in [20]. However, in this paper Walsh coefficients are used in ECOC design. A short summary of the prototype system for AU detection was first reported in [21], but used standard unweighted ECOC decoding. The use of bootstrapping and bias and variance analysis was applied to general (UCI) benchmark data sets in [22,23], but in this paper they are applied to weighted ECOC decoding for AU detection. In addition the weighted ECOC algorithm is derived from the first order Walsh coefficients in Section 3.4. A further contribution of the paper is to show that it is possible to generate individual AU probabilities by detecting AU groups using ECOC in Section 3.5.

The remainder of this paper is structured as follows: Section 2 reviews some of the algorithms that have been used in the field of facial expression recognition for performing feature extraction and feature selection, with particular reference to LBP and FCBF. Section 3 describes in some detail the method and rationale of weighted ECOC classification. A high-level description of the proposed approach to AU recognition is given in Section 4 and the results of performing experiments on a prototype of the main elements of the system are shown in Section 5. Finally some conclusions to be drawn from this work are discussed in Section 6.

2. Facial expression recognition

Given a raw input image, in the form of a 2-dimensional grid of grey-scale pixel values, a number of methods for extracting features that are suitable for facial expression recognition have been explored in the literature [7,24,25]. Useful surveys can be found in [26–28]. Many of the approaches to face expression recognition have been adapted from the field of face recognition, as the two problems have much in common. A method is described in [29] for both face and expression recognition which depends on mapping the high dimensional pixel feature space onto a low dimensional non-linear manifold. Other feature extraction methods that have been used to perform facial expression recognition include Gabor filters [8,2] and Haar features [30].

2.1. Local binary pattern features

The feature extraction method of local binary patterns (LBP) has been widely used in face recognition [31] and other facial analysis applications including facial expression recognition [see [32], for a survey]. The LBP operator is a powerful 2D texture descriptor that has the benefit of being somewhat insensitive to variations in the lighting and orientation of an image. As illustrated in Fig. 2, the algorithm associates each interior pixel of an intensity image with a binary code number in the range of 0–255. This code number is generated by taking the surrounding pixels and, working in a clockwise direction from the top left hand corner, assigning a bit value of 0 when the pixel intensity is less than that of the central pixel, and 1 otherwise. The concatenation of these bits produces an eight-digit binary code word which becomes the grey-scale value of the corresponding pixel in the transformed image. For sample points, that do not fall on the exact centre of a pixel, the grey-scale values can be obtained by interpolation from neighbouring pixels.

In order to use LBP codes as localised features for the purpose of image analysis, it is customary to cover an image with a set of overlapping or non-overlapping rectangular regions. Within each region a histogram of LBP codes is constructed and corresponding regions from two different images can be compared by using a measure such as chi-squared distance. In this paper we make use of an alternative method [33] in which the histograms from all regions are concatenated and each histogram bin constitutes a separate feature.

An important variation of the LBP operator is to treat all non-uniform patterns as indistinguishable. A uniform pattern is a pattern for which the circular neighbourhood of a pixel comprises at most one string of zeros and one string of ones. Such patterns occur more frequently and tend to be of greater significance than non-uniform patterns as they indicate straight lines, bright spots, dark spots, corners etc. For an 8-pixel neighbourhood there are 58 distinct uniform patterns so 59-bin histograms can be constructed, with the 59th bin holding all the non-uniform codes. Another modification to the basic LBP idea is to set the radius of the circle on which each LBP code is evaluated to a value greater than 1 pixel. This allows the LBP histograms to express larger scale textures in the image. A further possibility is to achieve finer resolution by increasing the number of bits per LBP code through the use of more than 8 (typically 16) circularly symmetric sampling points.

A number of researchers have used LBP-based methods to perform emotion recognition [3–5,34,35]. One problem with many applications of LBP is that a single LBP operator only represents texture information at one scale. The idea of multi-resolution LBP (MLBP) is to overcome this limitation by combining LBP operators at several scales. In [36] MLBP is applied to face recognition by selecting minimum-redundancy subsets of MLBP features and using them to discriminate between pairs of classes. Good face recognition performance is obtained in [9] by concatenating local histograms of MLBP features and using LDA within each region to reduce the dimensionality.

2.2. Feature selection

Having generated a large number of features from a face image it is necessary to reduce the dimensionality in order to improve classification accuracy and to reduce computational overheads. There are two ways of doing this, the first is to create new features by combining existing ones, for example by techniques such as PCA and LDA; the second is to select a subset of the existing