Pattern Recognition 47 (2014) 1271–1281

Contents lists available at ScienceDirect

Pattern Recognition

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

Facial expression recognition in dynamic sequences: An integrated approach

Hui Fang [a,d,*], Neil Mac Parthaláin [b], Andrew J. Aubrey [c], Gary K.L. Tam [a,c], Rita Borgo [a], Paul L. Rosin [c], Philip W. Grant [a], David Marshall [c], Min Chen [d]

[a] Computer Science Department, College of Science, Swansea University, United Kingdom
[b] Computer Science Department, Aberystwyth University, United Kingdom
[c] School of Computer Science & Informatics, Cardiff University, United Kingdom
[d] Oxford e-Research Centre, University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 24 April 2013
Received in revised form
16 August 2013
Accepted 19 September 2013
Available online 17 October 2013

Keywords:
Facial expression analysis
Dynamic feature extraction and visualisation

ABSTRACT

Automatic facial expression analysis aims to analyse human facial expressions and classify them into discrete categories. Methods based on existing work are reliant on extracting information from video sequences and employ either some form of subjective thresholding of dynamic information or attempt to identify the particular individual frames in which the expected behaviour occurs. These methods are inefficient as they require either additional subjective information, tedious manual work or fail to take advantage of the information contained in the dynamic signature from facial movements for the task of expression recognition.

In this paper, a novel framework is proposed for automatic facial expression analysis which extracts salient information from video sequences but does not rely on any subjective preprocessing or additional user-supplied information to select frames with peak expressions. The experimental framework demonstrates that the proposed method outperforms static expression recognition systems in terms of recognition rate. The approach does not rely on action units (AUs), and therefore, eliminates errors which are otherwise propagated to the final result due to incorrect initial identification of AUs. The proposed framework explores a parametric space of over 300 dimensions and is tested with six state-of-the-art machine learning techniques. Such robust and extensive experimentation provides an important foundation for the assessment of the performance for future work. A further contribution of the paper is offered in the form of a user study. This was conducted in order to investigate the correlation between human cognitive systems and the proposed framework for the understanding of human emotion classification and the reliability of public databases.

1. Introduction

Facial expression analysis has long been a research area of great interest. Indeed, work beginning as early as the nineteenth century [1] demonstrated that the analysis of facial expressions was of significance. The work in [2] was the first to formalise six different expressions that contained distinctive facial content. These six expressions were summarised as typical emotional displays of: *happiness*, *sadness*, *fear*, *disgust*, *surprise* and *anger*, and are now commonly known as the *basic emotions*. Until recently, the task of facial expression analysis has been a topic of research primarily associated with the field of psychology and much on the subject has been published in this area. However, interest broadened with the publication of the work in [3] which presented a preliminary investigation of the task of automatic facial expression analysis from a sequence of images. More recently, automatic facial expression analysis has attracted much attention particularly in the field of computer science. Some of the reasons for this are due to the advancements in related research sub-areas such as face detection [4], tracking and recognition [5], as well as new developments in the area of machine learning such as feature extraction and supervised learning [6,7].

Much of the recent work on facial expression analysis tended to focus on the ways of capturing the 'moment' or the point in time-series data (termed: *static expression recognition*) at which a particular facial expression begins to occur and when it ends. Previous approaches have mainly concentrated on attempting to capture expressions through either action units (AUs) [8,9] or from discrete frame extraction techniques [10]. All of these methods require either manual selection in order to determine where the

* Corresponding author. Tel.: +44 7886781646.
E-mail address: hui.fang@oerc.ox.ac.uk (H. Fang).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.023

particular behaviour occurs or the subjective imposition of thresholds. This means that any classification is highly dependent on the subjective information in the form of a threshold or other human-derived knowledge.

The approach proposed in this paper is formulated in order to tackle the aforementioned problems and to improve the performance of facial expression recognition by exploring dynamic signals. It offers a number of advantages over existing approaches: (a) the system does not require manual specification of the frame which shows peak expression; (b) the system uses the *dynamic information* of the facial features extracted from video sequences and outperforms techniques based on static images; and (c) it does not rely on the voting from groups of frames, where errors made earlier in the process are propagated leading to incorrect classification(s).

In addition to these advantages, a novel experimental evaluation presented in this paper offers a number of different perspectives for the task of facial expression analysis. For the learning of the expressions, six state-of-the-art machine learning methods are employed. Furthermore, an investigation of those sequences which are consistently mis-classified by the automatic methods is presented. This then forms the basis for a user study, which along with the use of visualisation tools offers an insight into the consistency of human perception and machine vision.

In summary, the contributions of the work are highlighted as follows:

- A novel automatic framework for the recognition of facial expressions using the dynamics of the sequences. Specific contributions include
  - The use of a group-wise registration algorithm to improve the robustness of tracking performance.
  - Construction of a parametric space of over 300 dimensions to represent the dynamics of facial expressions.
  - The use of six state-of-the-art machine learning methods for the automatic recognition task.
  - An objective comparison between the proposed system (which utilises dynamic information) and systems which utilise static apex images.
- Investigation of the correlation between human perception and machine vision for human emotion recognition.
  - The use of a visualisation technique for the analysis and initial understanding of facial feature data, and also for identifying outliers and noise in the data.
  - An intuitive user study to investigate the correlation between human perception and machine vision on facial expression recognition, and to assess the quality of a public dataset.

The remainder of this paper is structured as follows. Section 2 presents the background material for automatic facial expression analysis and provides an overview of current approaches. Section 3 describes the proposed approach (salient facial point tracking and feature extraction methods, and construction of dynamic signal parametric space) along with the automatic learning methods. Section 4 details the evaluation framework that is employed as well as the experimental setup and user survey. Finally, Section 5 concludes the paper along with some suggestions for further development.

2. Background

A system for automatic facial analysis may include many different aspects. Two of the most common are: (i) the automatic detection and classification of facial expressions – an area where much work has been carried out in the past [11,12], (ii) realistic facial expression synthesis in computer graphics [13], which is useful for studying the perception of expressions and also realistic computer animation; and (iii) expression analysis, important for affect recognition [14].

Typical facial expression recognition systems aim to classify an input facial image or video sequence into one of the six basic emotions mentioned previously. Facial expressions are formed through the movement of facial muscles, resulting in dynamic facial features such as the deformation of eyebrows, eyes, mouth and skin. Such changes can be captured and used in order to classify a given facial expression. In broad terms, there are two approaches (a) facial action unit (AU) based techniques and (b) content-based (non-AU) techniques; summarised in Sections 2.1 and 2.2 respectively.

2.1. Action unit based expression recognition

The facial action coding system (FACS) [2] is the most widely used method for describing the previously described facial movements. It defines 46 different action units (AUs) for the classification of non-rigid facial movements. This system forms the basis for many expression recognition systems [15–18].

In [19], several approaches that classify expressions are compared based on action unit classification accuracies. Some of these include Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Gabor filters and optical flow. It is claimed that by utilising local spatial features, better performance for expression analysis can be achieved. However, the use of techniques such as PCA destroys the underlying semantics of the local features making it more difficult to humanly interpret the results.

The work in [16] proposes the use of a rule-based system to learn facial actions by tracking salient points. Fifteen landmarks are tracked using a colour-based observation model via a particle filter algorithm applied to profile-view face images. A rule-based system is then implemented, by measuring the displacements of these salient points, in order to classify the sequences into discrete action units.

The relationship between action units using a dynamic Bayesian network is explored in [17]. The implicit assumption of this work is that the model is capable of representing the relationship amongst all AUs. Furthermore, it is claimed that AUs with weak intensity can be inferred robustly using other high-intensity AUs.

In more recent work [20], a system for emotion detection is proposed based on dynamic geometric features for AU activation detection which is then used within a hybrid SVM–HMM framework for emotion detection. The authors provide a robust analysis of their system and test the accuracy of its components on the MMI and *Cohn–Kanade* databases. However, emotion recognition performance is assessed using only the *Cohn–Kanade* database [21], so it is difficult to assess the generalisability of the approach.

2.2. Expression recognition without action units

For those methods which are not based on AUs, the two most common techniques for expression recognition utilise either static images that represent the apex of the expression [22] or the temporal facial dynamics [23].

In [24], grid nodes are tracked using a *Kanade–Lucas–Tomasi* tracker and the displacements of these nodes are extracted as features for training a support vector machine (SVM) in order to classify the six basic expressions. This work however only extracts geometric features after tracking.

Rather than utilising geometric features, the work in [22] implements a recognition system based on texture features called