Engineering Applications of Artificial Intelligence 38 (2015) 190–202

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Using multiple steerable filters and Bayesian regularization for facial expression recognition

Hela Mahersia*, Kamel Hamrouni

LR – Signal, Images et technologies de l'information, Ecole Nationale d'Ingénieurs de Tunis, Université Tunis El Manar, BP37, 1002 Tunis, Tunisie

A R T I C L E I N F O

Article history:
Received 16 June 2014
Received in revised form
28 September 2014
Accepted 2 November 2014
Available online 28 November 2014

Keywords:
Bayesian regularization neural network
Facial expression recognition
Steerable decomposition
Texture

A B S T R A C T

Facial expression recognition has recently become a challenging research area. Its applications include human–computer interfaces, human emotion analysis, and medical care and cure.

In this paper, we present a new challenging method to recognize seven universal emotional expressions, which are happiness, neutral, angry, disgust, sadness, fear and surprise. In our approach, we identify the user's facial expressions from the input images, using statistical features extracted from the steerable pyramid decomposition, and classified with a Bayesian regularization neural network. The evaluation of the proposed approach in terms of recognition accuracy is achieved using two universal databases, the Japanese Female Facial Expression database and the Cohn–Kanade facial expression database. The overall accuracy rate reaches 93.33% for the first database and is about 98.13% for the second one. These results show the effectiveness of the steerable proposed algorithm.

1. Introduction

According to numerous studies in psychology and psychiatry, faces are a very special source of inspiration. In particular, considerable attention has been paid to facial expressions, seen as an important way of information revealing the emotional state, intentions and beliefs in communication between persons. In this context, Mehrabian said in Mehrabian (1968) that the impact of facial expressions on the listener in a “face to face” conversation is much more important than the textual content of the message delivered. He shows that for a verbal message, 7% of the content is given by the meaning of words, 38% by the way according to which the words are spoken and 55% by facial expressions. This shows that facial expressions play a major role in human communications.

Facial expressions are generated by contractions of facial muscles, which results in temporally deformed features such as eye lids, eye brows, nose, lips and skin texture, often revealed by wrinkles and bulges. In the majority of research analyzing facial expressions, six universal emotions are considered (happiness, disgust, surprise, sadness, anger, and fear), to which is added the neutral expression (Ekman and Friesen, 1987) (Fig. 1).

Although facial expressions recognition has been studied by many researchers in the recent few years, the state of the art in this field is not very rich compared to facial recognition. A survey of the research made regarding facial expressions recognition can be found in Pantic and Rothkrantz (2000) and Fasela and Luettin (2003). Once a face is detected in the image, the corresponding region is extracted, and is usually normalized to have the same size. The next step is to extract facial features that describe the most appropriate representation of the face images for recognition. Mainly, two approaches can be distinguished: the template-based approach and the feature-based one.

The template-based approach aims to establish a model for each expression from the geometrical facial features and distances between them (including mouth, eyes, eyebrows and nose). The most commonly used methods are based on line detection (Gao et al., 2003), motion analysis from templates (Kotsia and Pitas, 2007), active appearance models (Abboud et al., 2004; Cheon and Kim, 2009) and facial action units (FAU) (Ekman and Friesen, 1987; Panatic and Pitas, 2006). The detected points can be extracted in 2D or 3D (Pantic and Rothkrantz, 2000; Dornaika and Raducanu, 2007; Venkatesh et al., 2009).

The feature-based methods use textural information as features for expression information extraction. Among the existing methods, we quote mainly the LBP (Zhao and Pietikainen, 2007; Shan et al., 2009), SIFT (Berretti et al., 2011), and Gabor filters (Lyons and Akamatsu, 1998; Lyons et al., 1999; Liu and Wang, 2006).

* Corresponding author: Tel.: +216 55 354 049.
E-mail addresses: helamahersia@yahoo.fr, hela.mahersia@enit.rnu.tn (H. Mahersia).

http://dx.doi.org/10.1016/j.engappai.2014.11.002
0952-1976/

Fig. 1. The seven universal emotions (from the Jaffe database Lyons et al., 1999).

This paper addresses these problems by proposing a new invariant textural descriptor based on steerable pyramids. A steerable pyramid is a method in which images are decomposed into a set of multi-scale, and multi-orientation image sub-bands, where the basis functions are directional derivative filters. In particular, the steerable filters introduced by Freeman and Adelson (1991) have attracted attention because they can analyze the contents of the image in many directions and at different levels of resolution. These filters are the heart of the steerable pyramid (Simoncelli et al., 1992, 1995; Mahersia et al., 2007; Mahersia, 2010).

The remainder of this paper is organized as follows. We first review the steerable pyramidal transform in Section 2, then, our work on facial expression recognition based on steerable filters is described in Section 3. In Section 4, we conduct some experiments in order to demonstrate the effectiveness of the proposed method. Finally, in Section 5 we draw the conclusions and outline future work to be carried out.

2. Steerable pyramid representation

In signal processing, a signal can be decomposed into sub-bands, such as by wavelet transform. The wavelet transform is widely used in many applications since the pyramid structure of wavelets responds well to a human visual system. However, the two major drawbacks of wavelets are the lack of translation invariance, especially in two-dimensional (2D) signals (Cheon and Kim, 2009) and the poor selectivity in orientation. To overcome this problem, the "steerable" pyramid wavelet, a class of arbitrary orientation filters generated by linear combination of a set of basis filters, has been proposed (Simoncelli et al., 1992, 1995).

The system diagram of a steerable pyramid for a single stage is shown in Fig. 2.

The pyramid is divided into two parts: analysis and synthesis. On the analysis part, the image is decomposed into low-pass and high-pass sub-bands, using steerable filters L_0 and H_0. The low-pass band continues to break down into a set of band-pass sub-bands $B_0, \ldots, B_K$ and lower low-pass sub-band L_1. The lower low-pass sub-band is sub-sampled by 2 along the two directions x and y. Repeating the shaded area provides the recursive structure.

Due to its invariant properties, the pyramid structure of the steerable wavelet was used in texture analysis (Montoya et al., 2007; Mahersia, 2010). It captures textures in both structural and random aspects (Portilla and Simoncelli, 2000; Mahersia et al., 2007).

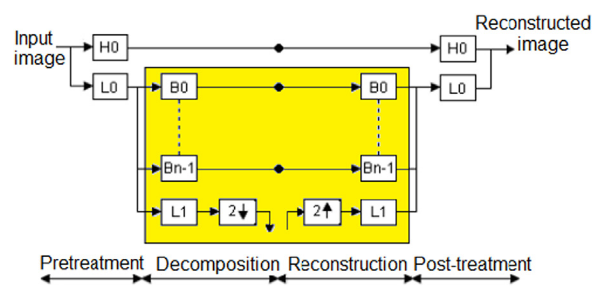

Fig. 2. First level of steerable pyramid decomposition system.

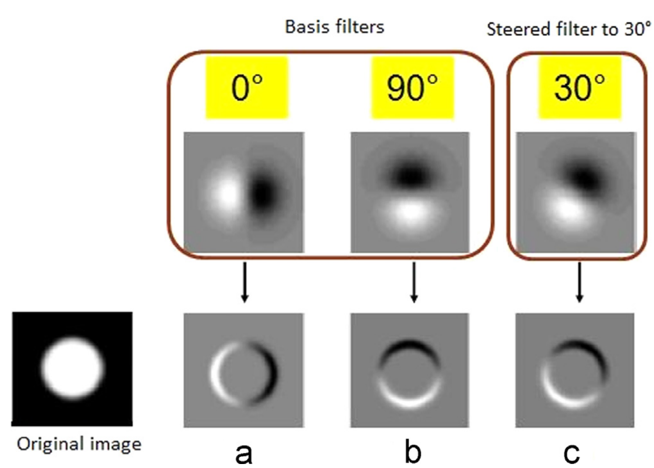

Fig. 3. Responses of the circular disk image convolved to the filter : (a) at 0°, (b) at 90° and (c) at 30°.

As shown in the work of Freeman and Adelson (1991), the derivatives of an image in any direction can be interpolated by several basis derivative functions. As shown in Ref. Montoya et al. (2007), any Kth-order directional derivative is a linear combination of $(K+1)$ Kth-order basis derivatives. For example, the response of the first-order steered filter in cartesian coordinates $f_1(x,y)$ to an arbitrary direction θ can be easily interpolated and synthesized by taking a linear combination of the basis filters $f_1^{0^\circ}$ and $f_1^{90^\circ}$ as shown