



Facial expression recognition using tracked facial actions: Classifier performance analysis

Fadi Dornaika^{a,b,*}, Abdelmalik Moujahid^a, Bogdan Raducanu^c

^a Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, San Sebastian, Spain

^b IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

^c Computer Vision center, Barcelona, Spain

ARTICLE INFO

Article history:

Received 19 August 2011

Received in revised form

2 September 2012

Accepted 3 September 2012

Available online 29 September 2012

Keywords:

Visual face tracking

3D deformable models

Facial actions

Dynamic facial expression recognition

Human–computer interaction

ABSTRACT

In this paper, we address the analysis and recognition of facial expressions in continuous videos. More precisely, we study classifiers performance that exploit head pose independent temporal facial action parameters. These are provided by an appearance-based 3D face tracker that simultaneously provides the 3D head pose and facial actions. The use of such tracker makes the recognition pose- and texture-independent. Two different schemes are studied. The first scheme adopts a dynamic time warping technique for recognizing expressions where training data are given by temporal signatures associated with different universal facial expressions. The second scheme models temporal signatures associated with facial actions with fixed length feature vectors (observations), and uses some machine learning algorithms in order to recognize the displayed expression. Experiments quantified the performance of different schemes. These were carried out on CMU video sequences and home-made video sequences. The results show that the use of dimension reduction techniques on the extracted time series can improve the classification performance. Moreover, these experiments show that the best recognition rate can be above 90%.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Overview

In recent times, there has been considerable technical progress within artificial intelligence in the field of computer vision to open the possibility of placing faces at the center of human–computer interaction (HCI). Facial expressions play an important role in recognition of human emotions. Psychologists postulate that facial expressions have a consistent and meaningful structure that can be backprojected in order to infer people inner affective state. Basic facial expressions typically recognized by psychologists are: happiness, sadness, fear, anger, disgust and surprise (Ekman, 1992). In the beginning, facial expression analysis was essentially a research topic for psychologists. However, recent progresses in image processing and pattern recognition have motivated significantly research works on automatic facial expression recognition (Fasel and Luetttin, 2003; Pantic and Patras, 2006; Yeasin et al., 2006). The question of how to further exploit the results of the recognized facial expression actually

motivates and fosters ongoing research in HCI, artificial intelligence and cognitive science. The field of ‘emotional machines’ (machines responsive to our emotions) is a vastly unexplored research domain with enormous potential.

A facial expression is formed by contracting or relaxing different facial muscles on human face which results in temporally deformed facial features like raising eyebrows and open mouth. The automated analysis of facial expressions is a challenging task because everyone’s face is unique and interpersonal differences exist in how people perform facial expressions. Numerous methodologies have been proposed to solve this problem (Bartlett et al., 2006; Cheon and Kim, 2009; Naghsh-Nilchi and Roshanzamir, 2006; Sebe et al., 2007; Xiao et al., 2011; Zeng et al., 2009; Zhang et al., 2008).

1.2. Related works

In the past, a lot of effort was dedicated to recognize facial expression in still images (static recognition). For this purpose, many techniques have been applied: neural networks (Tian et al., 2001), Gabor wavelets (Bartlett et al., 2006) and active appearance models (AAM) (Sung and Kim, 2009). A very important limitation to the static strategy for facial expression recognition is the fact that still images usually capture the apex of the expression, i.e., the instant at which the indicators of emotion are most

* Corresponding author at: Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Manuel Lardizabal, 1, 20018 San Sebastian, Spain.

E-mail address: fdornaika@hotmail.fr (F. Dornaika).

marked. Despite the fact that some of these techniques addressed non-apex expressions, their objective was to detect and recognize action units (e.g., Bartlett et al., 2006). In Zhang et al. (2012), the authors construct a sparse representation classifier (SRC). The effectiveness and robustness of the SRC method is investigated on clean and occluded facial expression images. Three typical facial features, i.e., the raw pixels, Gabor wavelets representation and local binary patterns (LBP) are extracted to evaluate the performance of the SRC method. In Moore and Bowden (2011), a sequential two stage approach is taken for pose classification and view dependent facial expression classification to investigate the effects of yaw variations from frontal to profile views. Local binary patterns (LBPs) and variations of LBPs as texture descriptors are investigated. Multi-class support vector machines are adopted to learn pose and pose dependent facial expression classifiers.

More recently, attention has been shifted particularly towards modeling dynamical facial expressions (Xiang et al., 2008; Robin et al., 2011). Recent research has shown that it is not just the particular facial expression, but also the associated dynamics that are important when attempting to decipher its meaning. The dynamics of facial expression can be defined as the intensity of the action units coupled with the timing of their formation. This is a very relevant observation, since for most of the communication act, people rather use 'subtle' facial expressions than showing deliberately exaggerated poses in order to convey their message. In Ambadar et al. (2005), the authors found that subtle expressions that were not identifiable in individual images suddenly became apparent when viewed in a video sequence.

Dynamical approaches can use shape deformations, texture dynamics (Yang et al., 2008) or a combination of them (Cheon and Kim, 2009). Dynamic classifiers try to capture the temporal pattern in the sequence of feature vectors related to each frame such as the hidden Markov models (HMMs) and dynamic Bayesian networks (Zhang and Ji, 2005). Cheon and Kim (2009) propose a dynamic recognition based on the differential active appearance model parameters. A sequence of input frames is fitted using the classical AAM, then a specific frame is selected as reference frame. The corresponding sequence of differential AAM parameters is recognized by computing the directed Hausdorff distance and the K nearest neighbor classifier. In Yeasin et al. (2006), a two-stage approach is used. Initially, a linear classification bank was applied and its output was fused to produce a characteristic signature for each universal facial expression. The signatures thus computed from the training data set were used to train discrete hidden Markov models to learn the underlying model for each facial expression. In Shan et al. (2006), the authors propose a Bayesian approach to modeling temporal transitions of facial expressions represented in a manifold. Xiang et al. (2008) propose a dynamic classifier that is based on building spatio-temporal model for each universal expression derived from Fourier transform. The recognition of unseen expressions uses Hausdorff distance in order to compute dissimilarity values for

classification. Dornaika and Raducanu (2007) propose a dynamic classifier that is based on an analysis-synthesis scheme exploiting learned predictive models given by second order Markov models. Local binary patterns have been used for facial expression recognition in Shan et al. (2009) and Zhao and Pietikinen (2007).

Wu et al. (2010) explore Gabor motion energy (GME) filters as a biologically inspired representation for dynamic facial expressions. They show that GME filters outperform the Gabor energy filters, particularly on difficult low intensity expression discrimination. Huang et al. (2011) combine some extracted facial feature sets using confidence level strategy. Noting that for different facial components, the contributions to the expression recognition are different, they propose a method for automatically learning different weights to components via the multiple kernel learning. Meng et al. (2011) use two types of descriptors motion history histogram (MHH) and histogram of local binary patterns (LBP). Based on these two basic types of descriptors, two new dynamic facial expression features are proposed. Moore et al. (2010) uses weak classifiers are formed by assembling edge fragments with chamfer scores. An ensemble framework is presented with all-pairs binary classifiers. An error correcting support vector machine (SVM) is utilized for final classification.

1.3. Paper contribution

Automatic facial expression recognition from video sequences is a very challenging task. Indeed, one has to use several modules in sequence: face detection, model fitting, 3D face tracking, face deformation tracking before applying a classifier that can infer the type of the displayed expression. Therefore, the problems of face detection, 3D face tracking, and facial action tracking are out of the scope of the paper. For the completeness of presentation, our face recognition system is depicted in Fig. 1. We stress the fact that the focus of the paper is on the third stage, namely the dynamic facial expression recognition. The majority of the proposed dynamic facial expression techniques assume high resolution frontal facial images. However, very few works have been done in order to recognize facial expression in the presence of head motion in 3D space. Although Moore and Bowden (2011) studied facial expression recognition under different poses, it is a static method that infers the expression from one single snapshot.

In this paper, we focus on the dynamic facial expression recognition in the presence of head motion. The recognition follows the extraction and tracking of facial actions using our 3D face and facial action tracking system (Dornaika and Davoine, 2006). Adopting such a 3D face tracker will overcome two main disadvantages associated with many existing dynamic recognition schemes. First, the expression recognition will not depend on the texture appearance, and hence more flexibility is gained in the sense that the learned models are independent from texture appearances and their changes (texture independence). This a clear

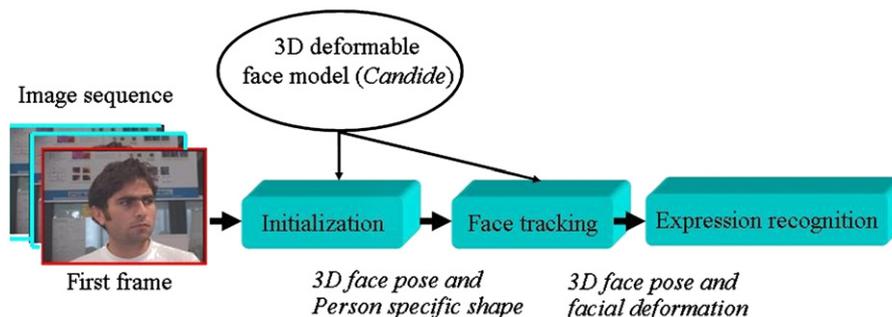


Fig. 1. Face recognition based on tracked facial deformation using the standard deformable face model Candide.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات