Chapter 14

Subtle Facial Expression Recognition in Still Images and Videos

Fadi Dornaika
University of the Basque Country, Spain & IKERBASQUE, Basque Foundation for Science, Spain

Bogdan Raducanu
Computer Vision Center, Spain

ABSTRACT

This chapter addresses the recognition of basic facial expressions. It has three main contributions. First, the authors introduce a view- and texture independent schemes that exploits facial action parameters estimated by an appearance-based 3D face tracker. They represent the learned facial actions associated with different facial expressions by time series. Two dynamic recognition schemes are proposed: (1) the first is based on conditional predictive models and on an analysis-synthesis scheme, and (2) the second is based on examples allowing straightforward use of machine learning approaches. Second, the authors propose an efficient recognition scheme based on the detection of keyframes in videos. Third, the authors compare the dynamic scheme with a static one based on analyzing individual snapshots and show that in general the former performs better than the latter. The authors then provide evaluations of performance using Linear Discriminant Analysis (LDA), Non parametric Discriminant Analysis (NDA), and Support Vector Machines (SVM).

INTRODUCTION

Many researchers in the engineering and computer science communities have been developing automatic ways for machines to recognize emotional expression, as a goal towards achieving human-machine intelligent interaction. Research on emotion classification utilizes pattern recognition approaches for recognizing emotions, using different modalities as inputs to the emotion recognition models. In the vision community, still images and videos depicting faces constitute the main channel that conveys human emotion. In the last two decades many automatic facial expression recognition methods have been proposed.

In the field of Human-Computer Interaction (HCI) computers will be enabled with perceptual capabilities in order to facilitate the communica-
tion protocols between people and machines. In other words, computers must use natural ways of communication people use in their everyday life: speech, hand and body gestures, facial expression. Recently, there is an increasing interest in non-verbal communication. Among it, affective computing plays a fundamental role: “computing that relates to, arises from, or deliberately influences emotions” (Picard, 1997). Indeed, the newest trend in computer systems aims at adapting them to the user’s needs and preferences through intelligent interfaces. From the AI perspective, undergoing research emphasize the strong relationship between cognition and emotion. A central part of the human-centred paradigm is represented by affective computing, i.e. computer systems able to recognize, interpret, and react accordingly to the affective phenomena perceived. The roots of affective computing can be found in psychology, which postulates that facial expressions have a consistent and meaningful structure that can be backprojected in order to infer people inner affective state (Ekman, 1993; Ekman & Davidson, 1994).

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 & Luettin, 2003; Kim et al., 2004; Yeasin et al., 2006). 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. In the past, a lot of effort was dedicated to recognize facial expression in still images. For this purpose, many techniques have been applied: neural networks (Tian et al, 2001), Gabor wavelets (Bartlett et al., 2004) and Active Appearance Models (AAM) (Sung et al., 2006).

A very important limitation to this strategy 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 marked. In their daily life, people seldom show apex of their facial expression during normal communication with their counterparts, unless for very specific cases and for very brief periods of time. In everyday life, the facial expressions observed are an interaction of emotional response and cultural convention. More recently, attention has been shifted particularly towards modeling dynamical facial expressions (Cohen et al., 2003; Shan et al., 2006; Yeasin et al., 2006).

Recent research has shown that it is not only the expression itself, but also its 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 (Zheng, 2000). In (Ambadar et al., 2005), the authors highlighted the fact that facial expressions are frequently subtle. They found that subtle expressions that were not identifiable in individual images suddenly became apparent when viewed in a video sequence. There is now a growing body of psychological research that argues that these dynamics are a critical factor for the interpretation of the observed behavior.

This chapter describes our recent works on facial expression recognition. We proposed two frameworks for facial expression recognition given natural head motion. These frameworks are based on an appearance-based 3D face tracker (Dornaika & Davoine, 2006). The proposed frameworks can be used for both kinds of facial expression: extreme posed expressions and subtle expressions. The first framework utilizes dynamic recognition schemes. The second framework utilizes static recognition schemes. While the dynamic recognition framework implicitly models subtle facial expressions, the second framework can perform the same task provided that the training examples (still images) include such occurrences. Within the first framework, we propose two different