Automatic Facial Expression Analysis

Huachun Tan
Tsinghua University, Beijing, China

Yujin Zhang
Tsinghua University, Beijing, China

INTRODUCTION

Facial expression analysis is an active area in human-computer interaction. Many techniques of facial expression analysis have been proposed that try to make the interaction tighter and more efficient.

The essence of facial expression analysis is to recognize facial actions or to perceive human emotion through the changes of the face surface. Generally, there are three main steps in analyzing facial expression. First, the face should be detected in the image or the first frame of image sequences. Second, the representation of facial expression should be determined, and the data related to facial expression should be extracted from the image or the following image sequences. Finally, a mechanism of classification should be devised to classify the facial expression data.

In this article, the techniques for automatic facial expression analysis will be discussed. The attempt is to classify various methods to some categories instead of giving an exhausted review.

The rest of this article is organized as follows. Background is presented briefly firstly. The techniques used in the three steps, which are detecting the face, representing the facial expression, and classifying the facial expression, are described respectively. Then some facial expression databases are discussed. The challenges and future trends to facial expression analysis are also presented. Finally, the conclusion is made.

BACKGROUND

During the past decade, the development of image analysis, object tracking, pattern recognition, computer vision, and computer hardware has brought facial expression into human-computer interaction as a new modality, and makes the interaction tighter and more efficient. Many systems for automatic facial expression have been developed since the pioneering work of Mase and Pentland (1991). Some surveys of automatic facial expression analysis (Fasel & Luettin, 2003; Pantic & Rothkrantz, 2000a) have also appeared.

Various applications using automatic facial expression analysis can be envisaged in the near future, fostering further interest in doing research in different areas (Fasel & Luettin, 2003). However, there are still many challenges to develop an ideal automatic facial expression analysis system.

DETECTING FACE

Before dealing with the information of facial expression, the face should be located in images or sequences. Given an arbitrary image, the goal of face detection is to determine whether or not there are faces in the image, and if present, return the location and extent of each face. Two good surveys of face detection have been published recently (Hjelmas, 2001; Yang, Kriegman, & Ahuja, 2002).

In most of the systems for facial expression analysis, it is assumed that only one face is contained in the image and the face is near the front view. Then, the main aim of this step is to locate the face and facial features.

In face detection, the input can be either a static image or an image sequence. Because the methods are totally different, we discuss them separately in the following paragraphs.

The techniques of face detection from static images can be classified into four categories (Yang et al., 2002), although some methods clearly overlap

the category boundaries. These four types of techniques are listed as follows:

- **Knowledge-Based Methods:** These rule-based methods encode human knowledge about what constitutes a typical face. Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization.
- **Feature-Invariant Approaches:** These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these features to locate faces. Usually, the facial features, such as the edge of the eye and mouth, texture, skin color, and the integration of these features, are used to locate faces. These methods are designed mainly for face localization.
- **Template-Matching Methods:** Several standard patterns of faces are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are computed for detection. Usually, predefined face templates and deformable templates are used. These methods have been used for both face localization and detection.
- **Appearance-Based Methods:** In contrast to template matching, models (or templates) are learned from a set of training images that should capture the representative variability of facial appearance. These learned models are then used for detection. Many learning models are studied, such as eigenface, the distribution-based method, neural networks, support vector machines, the Naïve Bayes classifier, the hidden Markov model, the information-theoretical approach, and so forth. These methods are designed mainly for face detection.

The face can also be detected by the motion information from image sequences. The approaches based on image sequences attempt to find the invariant features through face or head motion. They can be classified into two categories:

- **Accumulated Frame Difference:** In this type of approach, moving silhouettes (candidates) that include facial features, a face, or body parts are extracted by thresholding the accumulated frame difference. Then, some rules are set to measure the candidates. These approaches are straightforward and easy to realize. However, they are not robust enough to detect noise and insignificant motion.
- **Moving Image Contour:** In the approach, the motion is measured through the estimation of moving contours, such as optical flow. Compared to frame difference, results from moving contours are always more reliable, especially when the motion is insignificant (Hjelmas, 2001).

REPRESENTING FACIAL EXPRESSION

After determining the location of the face, the information of the facial expression can be extracted. In this step, the fundamental issue is how to represent the information of facial expression from a static image or an image sequence.

Benefiting from the development of image analysis, object and face tracking, and face recognition, many approaches have been proposed to represent the information of facial expression. These methods can be classified into different classes according to different criteria. Five kinds of approaches are discussed as follows.

According to the type of input, the approaches for representing facial expression can be classified into two categories:

- **Static-Image-Based Approaches:** The system analyzes the facial expression in static images. Typically, a neutral expression is needed to find the changes caused by facial expressions (Buciu, Kotropoulos, & Pitas, 2003; Chen & Huang, 2003; Gao, Leung, Hui, & Tanada, 2003; Pantic & Rothkrantz, 2000b, 2004).
- **Image-Sequence-Based Approaches:** The system attempts to extract the motion or changes of the face or facial features, and uses the spatial trajectories or spatiotemporal information to represent the facial expression information (Cohn, Sebe, Garg, Chen, & Huang, 2003; Essa & Pentland, 1997; Tian, Kanade, &