Chapter VI
Facial Expression Recognition

The facial expression has long been an interest for psychology, since Darwin published *The expression of Emotions in Man and Animals* (Darwin, C., 1899). Psychologists have studied to reveal the role and mechanism of the facial expression. One of the great discoveries of Darwin is that there exist prototypical facial expressions across multiple cultures on the earth, which provided the theoretical backgrounds for the vision researchers who tried to classify categories of the prototypical facial expressions from images. The representative 6 facial expressions are afraid, happy, sad, surprised, angry, and disgust (Mase, 1991; Yacoob and Davis, 1994). On the other hand, real facial expressions that we frequently meet in daily life consist of lots of distinct signals, which are subtly different. Further research on facial expressions required an object method to describe and measure the distinct activity of facial muscles. The *facial action coding system* (FACS), proposed by Hager and Ekman (1978), defines 46 distinct action units (AUs), each of which explains the activity of each distinct muscle or muscle group. The development of the objective description method also affected the vision researchers, who tried to detect the emergence of each AU (Tian et. al., 2001).

The general automated facial expression analysis system consists of three components: face acquisition, facial feature extraction, and facial expression classification (Fasel and Luettin, 2003). The facial expression acquisition includes face detection and facial image normalization to remove the effect of translation, scaling, and rotation. The facial feature extraction component extracts appropriate features for the facial expression recognition, where numerous techniques such as holistic image feature, geometric shape feature, optical flow, and wavelets are used. The recent trend is to adapt a model-based approach because
the face acquisition and facial feature extraction can be done relatively easy once the model is fitted to the input image successfully. The facial expression classification component classifies the extracted facial features into the one of the prototypical facial expressions or facial action units, where lots of classification algorithms are used such as neural network, support vector machines and Bayesian classifiers.

The model-based approaches can be divided along the face models into two groups: 2D face model group and 3D face model group. Among the two groups, many researchers have used 3D face models (Essa and Pentland, 1997a; Blanz and Vetter, 1999; Gokturk et. al., 2002). In general, using a 3D face model requires more computation time than traditional 2D based methods (Lyons et. al., 1999); however, it has several advantages. First, the problem of image distortion in the face image due to head pose change can be alleviated by using texture map (Blanz and Vetter, 1999). Second, a pose invariant 3D shape can be obtained, which is more adequate for natural facial expression recognition system than a 2D shape (Blanz and Vetter, 1999; Gokturk et. al., 2002). The approaches using a 2D shape (Zheng et. al., 2004, Michel and Kaliouby, 2003) always assume frontal view face images. Third, lighting problems can be alleviated by integrating the lighting model into a 3D face model fitting algorithm (Blanz and Vetter, 1999).

Although we can remove the effect of pose change or lighting conditions using the 3D face model, there exist many kinds of variations due to inter-person, facial expressions, verbal movements, and so on. These variations make the distributions of the facial expressions complicated and difficult to distinguish them. Some researchers have tried to separate these variations explicitly using bilinear model (Chuang et. al., 2002) or singular value decomposition (SVD) (Wang and Ahuja, 2003). Hence, such algorithms are inappropriate for explicit separation of the variations in on-line manner. Instead, other researchers used non-linear classification algorithms such as NN, or kernel methods to recognize facial expressions (Zheng et. al., 2004; Michel and Kaliouby, 2003). These approaches learn the complex manifolds itself. Another issue in facial expression recognition is to use dynamic features (Cohen et. al., 2003) rather than static features (Zheng et. al., 2004; Michel and Kaliouby, 2003). These approaches usually assume that they already know the neutral state and then compute the displacement of feature points or change of the parameters with respect to the neutral state. If it is possible to specify the image frame corresponding to neutral expression such information can be used to remove inter-person variation. However, specifying the exact neutral state of an unknown person is difficult.

This chapter is organized into two parts. The first part reviews some preliminary background such as the generalized discriminant analysis (GDA) (Baudat and Anouar, 2000), the bilinear models (BM) (Chuang et. al., 2002), and the relative expression image (REI) (Zhou and Lin, 2005). The GDA is an extension of the traditional discriminant analysis (Duda et. al., 2000) that are useful feature extraction algorithms for classification problems. The difference between two algorithms, LDA and GDA, is that the LDA is designed for linearly separable cases whereas the GDA is extended to handle non-linearly separable cases. Because the facial expression features have complex distributions in the feature space, the GDA is a good choice for the facial expression classification problems. The BM is a kind of image analysis technique especially useful when the images have two factors. For example, assume that we collected a set of face images for a specific person by changing the illumination direction and head pose. The collected face images have two factors that are illumination direction and head pose. When the collected face images are given, the BM can extract two parameter vectors corresponding to the illumination direction and
Related Content

Multiple Description Coding for Multipath Video Streaming
www.igi-global.com/chapter/multiple-description-coding-for-multipath-video-streaming/164631?camid=4v1a

Investigation of Human Monitoring Capabilities for Multiple Watch Windows
www.igi-global.com/article/investigation-of-human-monitoring-capabilities-for-multiple-watch-windows/177209?camid=4v1a

Beyond Communication and Control: Environmental Control and Mobility by Gaze
www.igi-global.com/chapter/beyond-communication-control/60037?camid=4v1a
Attention-Based Health Monitoring
www.igi-global.com/article/attention-based-health-monitoring/97702?camid=4v1a