Chapter 2

Spontaneous Facial Expression Analysis and Synthesis for Interactive Facial Animation

Yongmian Zhang
Rensselaer Polytechnic Institute, USA

Jixu Chen
Rensselaer Polytechnic Institute, USA

Yan Tong
GE Global Research, USA

Qiang Ji
Rensselaer Polytechnic Institute, USA

ABSTRACT

This chapter describes a probabilistic framework for faithful reproduction of spontaneous facial expressions on a synthetic face model in a real time interactive application. The framework consists of a coupled Bayesian network (BN) to unify the facial expression analysis and synthesis into one coherent structure. At the analysis end, we cast the facial action coding system (FACS) into a dynamic Bayesian network (DBN) to capture relationships between facial expressions and the facial motions as well as their uncertainties and dynamics. The observations fed into the DBN facial expression model are measurements of facial action units (AUs) generated by an AU model. Also implemented by a DBN, the AU model captures the rigid head movements and nonrigid facial muscular movements of a spontaneous facial expression. At the synthesizer, a static BN reconstructs the Facial Animation Parameters (FAPs) and their intensity through the top-down inference according to the current state of facial expression and pose information output by the analysis end. The two BNs are connected statically through a data stream link. The novelty of using the coupled BN brings about several benefits. First, a facial expression is inferred through both spatial and temporal inference so that the perceptual quality of animation is less affected by the misdetection of facial features. Second, more realistic looking facial expressions can be reproduced by modeling the dynamics of human expressions in facial expression analysis. Third, very low bitrate (9 bytes per frame) in data transmission can be achieved.

DOI: 10.4018/978-1-60960-024-2.ch002
INTRODUCTION

Facial expressions provide various types of messages for human communications. Facial expression synthesis is clearly of interest for many multimedia applications such as human-computer interaction, entertainment, virtual agents, interactive gaming, computer based learning, video teleconferences, and animation. To interactively synthesize the facial expressions of a live person, we need an automated facial expression analysis system, which can recognize the spontaneous facial expressions, by explicitly modeling their temporal behavior so that the various stages of the development of a human emotion can be interpreted by machine. However, extending the existing methods to spontaneous facial behavior analysis and synthesis is a non-trivial problem due to the following challenges.

1. Thousands of distinct nonrigid facial muscular movements related to facial actions have been observed so far (Scherer, 1982), and most of them differ subtly in a few facial features.
2. Compared to the highly controlled conditions of posed facial expressions, spontaneous facial expressions often co-occur with natural head movement when people communicate with others.
3. Unlike the posed facial expressions, most of the spontaneous facial expressions are activated without significant facial appearance changes. In addition, the spontaneous facial expression often has a slower onset phase and a slower offset phase compared to the posed facial expression. (Cohn, 2004)
4. The spontaneous facial expression may have multiple facial expressions often occurring sequentially without always following a neutral-expression-neutral temporal pattern (Pantic, 2006) as for the posed facial expressions.

Since the MPEG-4 visual standard (MPEG4, 1998) will have a crucial role in forthcoming multimedia applications, the facial expression synthesis has gained much interest within the MPEG-4 framework. The MPEG-4 visual standard specifies a set of facial definition parameters (FDPs) and facial animation parameters (FAPs). The FAPs are used to characterize the movements of facial features defined over jaw, lips, eyes, mouth, nose, cheek, which are adequate to define the measurement of muscular actions relevant to AUs. Moreover, the FAPs can be placed on any synthetic facial model in a consistent manner with little influence by the inter-personal variations. For animating a person in a remote end, the FDPs are normally transmitted once per session and then followed by a stream of compressed FAPs. The animation of a virtual face is achieved by first transmitting the coded FAPs and then resynthesizing on the client-side. To accommodate very low bandwidth constraint, the FAPs must be compressed. Despite significant progress, the current techniques in facial expression synthesis face several issues that still need to be resolved.

1. Although the discrete cosine transform (DCT) technique can achieve a high FAP compression, the DCT involves a large coding delay (temporal latency) that makes it unsuitable for real time interactive applications. The principal component analysis (PCA) is able to achieve a high FAP compression for intraframe coding, however, it compromises the reconstruction accuracy.
2. An automated video analyzer may often mis-detect some facial features. Consequently, this may create animation artifacts on the facial model, which affects the perceptual quality of facial animation.
3. The intensity of facial expressions reveals the emotional evolution. It is difficult for machine to extract the subtle variation of facial features. Consequently, a dynamic