Abstract—This paper explores the use of multisensory information fusion technique with Dynamic Bayesian networks (DBNs) for

modeling and understanding the temporal behaviors of facial expressions in image sequences. Our facial feature detection and tracking

based on active IR illumination provides reliable visual information under variable lighting and head motion. Our approach to facial

expression recognition lies in the proposed dynamic and probabilistic framework based on combining DBNs with Ekman’s Facial Action

Coding System (FACS) for systematically modeling the dynamic and stochastic behaviors of spontaneous facial expressions. The

framework not only provides a coherent and unified hierarchical probabilistic framework to represent spatial and temporal information

related to facial expressions, but also allows us to actively select the most informative visual cues from the available information sources

to minimize the ambiguity in recognition. The recognition of facial expressions is accomplished by fusing not only from the current visual

observations, but also from the previous visual evidences. Consequently, the recognition becomes more robust and accurate through

explicitly modeling temporal behavior of facial expression. In this paper, we present the theoretical foundation underlying the proposed

probabilistic and dynamic framework for facial expression modeling and understanding. Experimental results demonstrate that our

approach can accurately and robustly recognize spontaneous facial expressions from an image sequence under different conditions.

Index Terms—Facial expression analysis, dynamic Bayesian networks, visual information fusion, active sensing.

226

***Abstract*—The tracking and recognition of facial activities**

**from images or videos have attracted great attention in computer**

**vision field. Facial activities are characterized by three levels.**

**First, in the bottom level, facial feature points around each**

**facial component, i.e., eyebrow, mouth, etc., capture the detailed**

**face shape information. Second, in the middle level, facial action**

**units, defined in the facial action coding system, represent the**

**contraction of a specific set of facial muscles, i.e., lid tightener,**

**eyebrow raiser, etc. Finally, in the top level, six prototypical facial**

**expressions represent the global facial muscle movement and are**

**commonly used to describe the human emotion states. In contrast**

**to the mainstream approaches, which usually only focus on one or**

**two levels of facial activities, and track (or recognize) them separately,**

**this paper introduces a unified probabilistic framework**

**based on the dynamic Bayesian network to simultaneously and**

**coherently represent the facial evolvement in different levels, their**

**interactions and their observations. Advanced machine learning**

**methods are introduced to learn the model based on both training**

**data and subjective prior knowledge. Given the model and the**

**measurements of facial motions, all three levels of facial activities**

**are simultaneously recognized through a probabilistic inference.**

**Extensive experiments are performed to illustrate the feasibility**

**and effectiveness of the proposed model on all three level facial**

**activities.**

***Index Terms*—Bayesian network, expression recognition, facial**

**action unit recognition, facial feature tracking, simultaneous**

**tracking and recognition.**

236

***Abstract—*Automatic analysis of human facial expression is a**

**challenging problem with many applications. Most of the existing**

**automated systems for facial expression analysis attempt to recognize**

**a few prototypic emotional expressions, such as anger and**

**happiness. Instead of representing another approach to machine**

**analysis of prototypic facial expressions of emotion, the method**

**presented in this paper attempts to handle a large range of human**

**facial behavior by recognizing facial muscle actions that produce**

**expressions. Virtually all of the existing vision systems for facial**

**muscle action detection deal only with frontal-view face images and**

**cannot handle temporal dynamics of facial actions. In this paper,**

**we present a system for automatic recognition of facial action units**

**(AUs) and their temporal models fromlong, profile-view face image**

**sequences. We exploit particle filtering to track 15 facial points in**

**an input face-profile sequence, and we introduce facial-action-dynamics**

**recognition from continuous video input using temporal**

**rules. The algorithm performs both automatic segmentation of an**

**input video into facial expressions pictured and recognition of temporal**

**segments (i.e., onset, apex, offset) of 27 AUs occurring alone**

**or in a combination in the input face-profile video. A recognition**

**rate of 87% is achieved.**

***Index Terms—*Computer vision, facial action units, facial expression**

**analysis, facial expression dynamics analysis, particle filtering,**

**rule-based reasoning, spatial reasoning, temporal reasoning.**

217