

Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition

Yuan An

Today, I read a review paper *Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition* by Sariyanidi [2]. The paper reviews the progress across a range of affect recognition applications to shed light on these fundamental questions. And it analyses the state-of-the-art solutions by decomposing their pipelines into fundamental components, namely face registration, representation, dimensionality reduction and recognition. Then the role of these components is discussed and the models and new trends that are followed in their design are highlighted. Moreover, the authors provide a comprehensive analysis of facial representations by uncovering their advantages and limitations, elaborate on the type of information they encode and discuss how they deal with the key challenges of illumination variations, registration errors, head-pose variations, occlusions, and identity bias.

Introduction

The production, perception and interpretation of facial expressions have been analyzed for a long time across various disciplines such as biology, psychology, neuroscience, sociology and computer science. The authors tell us the difference of between cognitive sciences and computer vision and machine learning in application. While cognitive sciences provide guidance to the question *what* to encode in facial representations, computer vision and machine learning influence *how* to encode this information. The appropriate cues to interpret facial expressions and how to encode them remain open questions.

Facial representation can be divided into spatial or spatio-temporal. Spatial representations encode image sequences frame-by-frame, whereas spatio-temporal representations consider a neighborhood of frames. Another classification is based on the type of information encoded in space: appearance or shape. Appearance representations use textural information by considering the intensity values of the pixels, whereas shape representations ignore texture and describe shape explicitly.

The main challenges in automatic affect recognition are head-pose variations, illumination variations, registration errors, occlusions and identity bias.

Advances in the field, and the transition from controlled

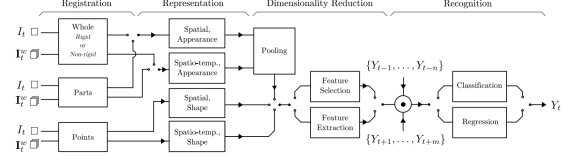


Figure 1. The proposed conceptual framework to be used for the analysis and comparison of facial affect recognition systems. The input is a single image (I_t) for spatial representations or a set of frames (I_t^ω) within a temporal window ω for spatio-temporal representations. The system output Y_t is discrete if it is obtained through classification or continuous if obtained through regression. The recognition process can incorporate previous (Y_{t-1}, \dots, Y_{t-n}) and/or subsequent (Y_{t+1}, \dots, Y_{t+m}) system output(s).

to naturalistic settings have been the focus of a number of survey papers. Zeng *et al.* [3] focused on automatic affect recognition using visual and auditory modalities. Gunes and Schuller [1] highlighted the continuity aspect for affect recognition both in terms of input and system output.

Yet no survey has analyzed systems by isolating their fundamental components (see Figure 1) and discussing how each component addresses the above-mentioned challenges in facial affect recognition.

In the paper, the authors break down facial affect recognition systems into their fundamental components (see Figure 1): facial registration, representation, dimensionality reduction and recognition. Then discuss the role of each component in dealing with the challenges in affect recognition. And they analyze facial representations in detail by discussing their advantages and limitations, the type of information they encode, their ability to recognize subtle expressions, their dimensionality and computational complexity.

Affect models and recognition

Affect recognition systems aim at recognizing the appearance of facial actions or the emotions conveyed by the actions. The former set of systems usually rely on the Facial Action Coding System (FACS). FACS consists of facial Action Units (AUs), which are codes that describe certain facial configurations.

The systems that recognize emotions consider basic or

non-basic emotions. Basic emotions refer to the affect model developed by Ekman and his colleagues, who argued that the production and interpretation of certain expressions are hard-wired in our brain and are recognized universally. The emotions conveyed by these expressions are modelled with six classes: happiness, sadness, surprise, fear, anger and disgust. Basic emotions are believed to be limited in their ability to represent the broad range of everyday emotions [1].

References

- [1] H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, 2013. [1](#), [2](#)
- [2] E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133, 2015. [1](#)
- [3] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009. [1](#)