34. MANISH JAYAN

ISLA: Temporal Segmentation and Labeling for Audio-Visual Emotion Recognition

Abstract

Emotion is an essential part of human interaction. Automatic emotion recognition can greatly benefit human-centered interactive technology, since extracted emotion can be used to understand and respond to user needs. However, real-world emotion recognition faces a central challenge when a user is speaking: facial movements due to speech are often confused with facial movements related to emotion. Recent studies have found that the use of phonetic information can reduce speech-related variability in the lower face region. However, methods to differentiate upper face movements due to emotion and due to speech have been under explored. This gap leads to the proposal of the Informed Segmentation and Labeling Approach (ISLA). ISLA uses speech signals that alter the dynamics of the lower and upper face regions.This demonstrated how pitch can be used to improve estimates of emotion from the upper face, and how this estimate can be combined with emotion estimates from the lower face and speech in a multimodal classication system. Human emotion classication results on the IEMOCAP and SAVEE datasets show that ISLA improves overall classication performance. Aim is to demonstrate how emotion estimates from different modalities correlate with each other, providing insights into the differences between posed and spontaneous expressions.