The ability of a computer to detect and appropriately respond to changes in a user’s affective state has significant implications to Human–Computer Interaction (HCI). To more accurately simulate the human ability to assess affects through multi-sensory data, automatic affect recognition should also make use of multimodal data. In this paper, we present our efforts toward audio–visual affect recognition. Based on psychological research, we have chosen affect categories based on an activation–evaluation space which is robust in capturing significant aspects of emotion. [continued in the next page]
We apply the Fisher boosting learning algorithm which can build a strong classifier by combining a small set of weak classification functions. Our experimental results show with 30 Fisher features, the testing error rates of our bimodal affect recognition is about 0.16 on the evaluation axis and 0.13 on the activation axis.