Mutual Information-Based Selection of Audiovisual Affective Features to Predict Instantaneous Emotional State

Sudipta Paul†, Nurani Saoda†, S M Mahbubur Rahman†, *Member*, *IEEE*, Dimitrios Hatzinakos‡, *Fellow*, *IEEE*
†Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
‡Identity, Privacy and Security Institute, Department of Electrical and Computer Engineering
University of Toronto, Toronto, ON, Canada, M5S 2E4
E-mail: {sudiptapaul.paul, saodacynthia}@gmail.com, mahbubur@eee.buet.ac.bd, dimitris@comm.utoronto.ca

***Abstract*—Automatic prediction of continuous level emotional state requires selection of suitable affective features to develop a regression system based on supervised machine learning. This paper investigates the performance of low-level dynamic features for predicting two common dimensions of emotional state, namely, valence and arousal instantaneously. Low-complexity features are extracted from audio and visual modalities independently and fused in the feature level. Features with minimum redundancy and maximum relevancy are chosen by using the mutual information-based selection process. The performance of frame-by-frame prediction of emotional state using the moderate length features as proposed in this paper is evaluated on spontaneous and naturalistic human-human conversation of SEMAINE database. Experimental results show that the proposed features selected by mutual information can be used for instantaneous prediction of emotional state with an accuracy higher than traditional audio or visual features that are used for affective computation.**

I. Introduction

Human can easily understand the instantaneous affective information conveyed by speakers during human-human conversation from multimodal cues. Automatic retrieval of affective information helps to develop artificial listener agents and emerging user-oriented technologies. For example, machines will be able to provide flexible performance under uncertain conditions, movie directors can change affective content of a video aiming at certain viewer groups, emotional state of drivers can be monitored to engage safety measures, and impact of television commercials or programs can be measured by judging the emotional state of the viewers remotely. Thus, automatic prediction of instantaneous affective state is becoming increasingly important in the recent years [1].

A. Related Works

Analysis of affective content is an interdisciplinary field involving research areas that includes computer vision, speech analysis, and psychology. To relate between measurable low-level features with corresponding affective state, certain models of emotion are required. Psychologists have used two major approaches, viz., categorical and dimensional to quantify the emotional states [2]. According to the categorical approach, the model of emotion was defined by Ekman [3], who grouped emotion into six basic categories including the happiness, sadness, anger, disgust, fear, and surprise. Situation arises where a small number of discrete categories may not reflect the complexity of emotional states. In this context, continuous emotional model reflect more subtle and context specific emotions avoiding boundaries. As a result, research in area of affective computing is shifting from categorical approach to dimensional approach. Wundt [2] introduced dimensional approach, wherein the emotion is divided into 3D continuous space - valence, arousal, and dominance. Valence represents the degree of pleasure, ranging from pleasant to unpleasant feelings. Arousal illustrates the activation level ranging from global feeling of dynamism to lethargy of an individual. Dominance characterizes the range of emotion from controlling sentiment to the controlled or submissive feelings. Dietz and Lang [4] have shown that the effect of the dominance dimension becomes visible only at points with

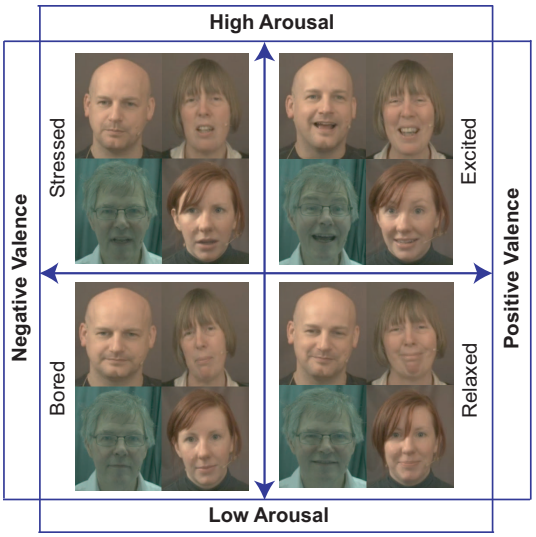

Fig. 1. Examples of facial appearances in the valence-arousal plane showing typical emotional states.

978-1-5090-4090-2/16/$31.00 ©2016 IEEE

ISBN 978-1-5090-4089-6