A COMPREHENSIVE BIOINFORMATION SYSTEM OF HUMAN EMOTION RECOGNITION

Siarhei Sadau, Igor Kheidorov

Department of Radio Physics and Computer Technologies, Belarusian State University, Belarus
seregasadov@gmail.com

Most of the known computer systems which determine the emotional state of a person are based on obtaining data from a facial image in a video sequence. A more promising direction is based on the analysis of multiple streams: facial images, gestures, postures, speech and biometrical data. The proposed system allows obtaining information by analyzing facial expressions, speech and biometrical data of the object in real time.
The system (figure 1) consists of four main modules: data acquisition module, data processing module, classification module, and a data output module. Let us consider in more detail each component.

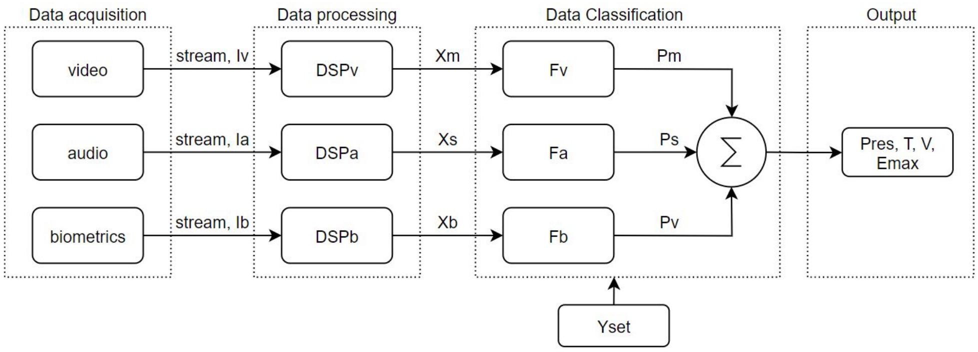

Fig 1 A model of the system

Data acquisition module
From the analyzed object the data is fed to the input of video, audio, and biometric channels. Each channel carries out a set of unique operations on the input data. The video channel performs the selection and capture of the face in the video stream, the segmentation of key areas that demonstrate muscle movement. The audio channel carries out the primary selection of the useful signal (human speech) in the audio stream. The biometric channel supplies a stream of data on skin temperature, blood pressure, and heart rate.
As a rule, the analyzed object locates in a noisy environment, therefore there is a non-zero probability of obtaining poor-quality data for each channel, up to the complete impossibility of using them. To take this feature into account, each channel must calculate its own coefficient of content relevance I_{ch} and transfer it to the processing module.
Data processing module
Such a module receives at its input previously prepared information for each of the channels and outputs normalized vectors X_{set} of data attributes at the output. Such vectors resulting from digital signal processing are unique to each non-verbal component. For example, for the mimic component, the vector contains information about the change in the location of the supporting segments on the face, which characterize the dynamics of contraction of the facial muscles. For speech - the volume, pitch and frequency response of speech.
Additional Y_{set} features to X_{set} feature vectors take into account the age, educational characteristics of a person, his gender, nationality, etc. since recent studies [1] cast doubt on the thesis of the invariance of the expression of emotions by different people.
Classification Module
The main output parameter, the P_{res} vector, reflects the probability that the non-verbal component corresponds to a certain type of emotion from a given set of considered emotions E_{set}. Functional dependencies F_{ch}: $X_{set} \rightarrow P_{ch}$, where *ch* is the corresponding non-verbal channel, the values of the feature vectors are mapped onto probability vectors P_{ch}. For such a functional display, classifiers [2], unique for each channel, which are based on a training data set, are used. Further, the output of the classifiers, P_{ch} vectors, must be normalized and summed to obtain the integral vector P_{res}.
Data output module
As a result, we can determine the parameter $E_{res} \in E_{Set}$ corresponding to the dominant type of emotion highlighted in the general set of possible emotional states
The output parameterization used also allows obtaining secondary output data - valency and tonality.

[1] Gendron M, Roberson D, van der Vyver JM, Barrett LF // Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture, Affective Science Institute, - 2014.
[2] Soroka, A. M. Recognition of emotions in speech based on wavelet analysis with adaptable basic functions / A.M. Soroka A.V. Semenchenko, I.E. Kheidorov // Information systems and technologies: materials of the international congress on computer science.- Minsk: BSU, 2016, c.646-651.