Using Remote Heart Rate Measurement for Affect Detection

Hamed Monkaresi, M. Sazzad Hussain, Rafael A. Calvo

School of Electrical and Information Engineering,
The University of Sydney, Sydney, NSW, 2006, Australia
{hamed.monkaresi, sazzad.hussain, rafael.calvo}@sydney.edu.au

Abstract

Current research suggests that using multiple modalities in affect detection techniques can improve their accuracy. Combining facial expression and physiological signals is one of the most common approaches in multimodal affect detection. Several methods and devices have been invented for measuring physiological signals in a simple way and have been used widely in affective computing applications. Out of the various approaches, contact-less sensors which can measure physiological signals remotely are more desirable to be used in everyday life and applications. In this paper we proposed a new fusion model for affect detection which extracts facial expression features and heart rate using a single video recording sensor. To our knowledge this is the first attempt to use remote physiological signal sensor for affect detection. The results suggest that fusing these heart features can improve the accuracy of affect detection systems.

Introduction

Several methods, techniques and devices have been proposed in the past for affect detection. Some of them relied on single modalities like facial expression, voice and physiological signals which were successful for detecting even complex affective states (Calvo & D'Mello, 2010). However, multimodal affect detection techniques are becoming increasingly popular due to their better reliability and performance in detecting complex affective states (D'Mello & Kory, 2012; Nicolaou, Gunes, & Pantic, 2011). Naturally, humans use several modalities when they are interacting with each other. Each modality (face, voice, gesture, physiology, etc.) has a unique characteristic of an affective state and considering more modalities can increase reliability and accuracy of affect interpretation.

Physiology is one of the prominent modalities that has been used for affect detection because of its suitability for reflecting inner feeling and robust against deceptive behavior. It has also been used in multimodal affect detection approaches (Hussain, Calvo, & Pour, 2011; Soleymani, Pantic, & Pun, 2012).

Normally, physiological sensors need to be attached to the human body which might be intrusive and make the application hard to adopt. Wearable sensors and devices were proposed to reduce the hardships of setting up the traditional sensors. Among the current methods of measuring physiological signals, contact-less and remote methods are more desirable. These methods are easy to adopt and cheaper than traditional devices (Poh, McDuff, & Picard, 2010). A remote, contactless sensor could monitor several subjects at the same time.

In this paper we have introduced a new method to measure heart rate (HR) remotely and use it for affect detection in combination with facial expression features. A bimodal system is proposed which used only one sensor: a camera. A dynamic approach has been used to extract facial expression features based on local binary patterns in three orthogonal planes (LBPTOP) (Monkaresi, Hussain, & Calvo, 2012). As for the second modality, a variation of Poh et al. (Poh et al., 2010) method has been used to extract HR features. Then a fusion model has been utilized to classify affective states using these two modalities.

Related work

Multimodal affect detection

Information fusion is likely to improve detection accuracy by integrating features from multiple modalities (D. L. Hall & Llinas, 1997). Machine learning techniques are used for classifying affective states using the merged feature set or by combining decisions from modalities. Substantial progress in multimodal affect detection has been reported in the literature (Calvo & D'Mello, 2010), however only a handful of studies have considered physiological channels with other modalities (e.g. speech, facial videos). Moreover, fusing multimodal data for affect detection that can be practically implemented and generalized with good accuracy can be challenging (D'Mello & Kory, 2012), particularly for naturalistic scenarios.

Some emotions are better interpreted from facial expressions while others from speech, so researches (Busso et al., 2004) have proposed methods to combine the two in order to take advantage of their complimentary relationship. Gunes and Piccardi (Gunes & Piccardi, 2007) have used face and body to recognize emotion and their approach achieved better recognition accuracy compared to the face modality. Approaches using physiological signals with other modalities have also been reported. Kim and Andre (Kim & Andre, 2006) proposed the integration of physiological and speech signals and Bailenson et al. (Bailenson et al., 2008) used facial video and physiology for emotion recognition.

2.2. Remote measuring of physiological signal

Three main categories have been defined for remote measuring of vital signs (i.e. heart beat). The first category uses microwave Doppler radar (Li, Cummings, & Lam, 2009). Using microwave Doppler for detecting the heartbeats and breathing was the earliest remote contactless system developed in the 1980s. The second category is based on thermal imaging (Fei & Pavlidis, 2010), where remote HR detection is performed by analyzing skin temperature modulation.

Compared to other two categories, video based techniques (Poh et al., 2010; Verkruysse, Svaasand, & Nelson, 2008) are considered cheaper and easier to adopt. Most of these attempts use the photoplethysmography (PPG) methodology to detect cardiovascular blood volume pulse. The PPG method was introduced in 1937 (Hertzman & Spealman, 1937) and typically measures the light reflection from the skin tissue to detect some physiological signals like oxygen saturation (pulse oxymetry), heart rate, blood pressure, etc. It works based on this principle that blood generally absorbs light more compare to other tissue. Accordingly after each heart pulse, fresh blood is pumped to the skin and the capability of light absorption will be increased. Detectable changes in light absorption caused by heart pulses (pulsatile) is very small compare to the absorption due to non-pulsatile arterial blood. These small changes should be detected and amplified to measure heart rate signal.

Typically PPG uses a dedicated light source emits the light (e.g., infra-red wavelength) into the skin and measures the transmission or reflectance using another sensor. Verkruysse et al. (2008), for the first time, have shown that PPG signals could be detected remotely on human face with a normal ambient light (as the only illumination source) using an ordinary digital camera. This attempt opened a door to the lots of possible applications in remote physiological sensing research area. Later, an improvement on their work has been introduced by Poh et al. (Poh et al., 2010), using Independent Component Analysis (ICA) method. They have compared their algorithm with a BVP sensor and achieved a Pearson correlation coefficient of 0.98 for detecting HR at rest (Poh et al., 2010). However, Monkaresi et al. (Monkaresi et al. in press) demonstrated that their method is not as accurate for more naturalistic scenarios, such as HCI applications. They also proposed a machine learning approach to improve Poh et al.'s method for detecting HR during naturalistic HCI. According to their report, their proposed method improved the accuracy of HR detection during naturalistic human computer interaction by decreasing the mean squared error from 43.76 beats per minute (bpm) to 3.64 (bpm). In this paper, we have applied Monkaresi et al.'s implementation for extracting HR features.

3. Material and Data Collection

3.1. Participants

For this study, 23 undergraduate and postgraduate engineering students from the University of Sydney were recruited for the data collection. The participants' aged from 20 to 60 years (M=34 years, SD=11) and there were 14 males and 9 females. There were 5 Asians and 17 Caucasian and 6 participants wore eyeglasses. The study was approved by the University of Sydney's Human Ethics Research Committee prior to data collection.

There was a synchronization problem between video segments and self-reports for one participant. This participant was ignored for feature extraction and affect classification. Two participants were too close to the screen while viewing the stimuli and facial features could not be extracted. The eye-related features could not be extracted for three participants due to occlusion caused by eyeglasses. Therefore, the classification results are reported based on recorded data from 17 participants.

3.2. Procedure

A total of 60 images were used, each presented for 10 seconds, and followed by a 10 seconds pause between images for annotation (self reporting). The images were selected from the International Affective Picture System (IAPS) database (P. J. Lang, Bradley, & Cuthbert, 2008). It provides a set of emotional stimuli to trigger emotion and attention. Each picture in this database was rated by approximately 100 participants in terms of three dimensions: valence, arousal and dominant. A 1-9 rating scale were used for each dimension. It also provides gender specific ratings. Different patterns in rating are observed for each gender that shows the importance of considering "gender" during designing the emotion acquisition protocol. We used these rating to categorize two gender specific groups for our experiment. Each gender specific group was divided into four categories based on arousal and valance normative ratings. Fifteen images were selected for each category.

Two types of labels were considered for each video segment: *normative-rating* and *self-reports*. The normative-ratings were from the IAPS dataset. In addition, the self-reported ratings were collected during the experiment concurrently. After viewing each image, the participants were asked to fill-out a questionnaire in order to indicate their degree of valence and arousal. The Self-Assessment Manikin (SAM) protocol (P. Lang & Bradley, 1997) was used for this reason. These annotations were used as the ground truth for evaluating our proposed system.

3.3. Sensors and Experiment setup

The experiment was conducted indoors with a varying amount of ambient sunlight entering through windows in combination with normal artificial fluorescent light. Participants were asked to sit in front of a computer and interact normally while their video was recorded by a Microsoft Kinect sensor (PC version). All videos were recorded in color (24-bit RGB with 3 channels, 8 bits/channel) at 30 frames per seconds (fps) with pixel resolution of 640×480 pixels and saved in AVI format. The ECG was recorded using a BIOPAC MP150 system with AcqKnowledge (v. 3.8.2) software. The acquisition sampling rate was 250 Hz. Three electrodes were placed on participant's body to record ECG signals: two electrodes were placed on their arms and the ground electrode was placed on their ankle. The ECG signal was used for evaluating the video-based HR detection method.

4. Method

4.1 LBPTOP

The LBP method detects the local-patterns existing in an image. This method is proposed by Ojala et al. (Ojala, Pietikäinen, & Harwood, 1996) for describing texture images. By applying LBP operator on all pixels of an image (or a sub-region of an image) and computing the distribution of local-patterns, a unique histogram could be extracted which describes the occurrence of each specific local-patterns throughout that image. This histogram is a powerful identifier for each image and has been shown good performance in several pattern recognition applications. Having *P* neighbourhood pixels could describe 2^P distinguishable local-patterns. The size of each local-pattern could be defined by number of neighbourhood pixels (*P*) and the radii (*R*). The radii specify the distance between neighbourhood pixels and the centre point. The ideal values for *R* and *P* depend on the application domain and the characteristics of the image.

One of the successful extensions of LBP for detecting facial expressions has been introduced by (Zhao & Pietikäinen, 2007). To extract LBP features from a video segment, they have divided a video segment into three sets of orthogonal planes. A video segment could be considered as a sequence of static images (XY planes) in the time axis. In another perspective, we can analyse a video segment as a stack of XT planes in the Y axis and a stack of YT planes in the X axis. The LBPTOP features are extracted by calculating LBP features from these three orthogonal planes.

We have used a recent implementation of LBPTOP for affect detection (Monkaresi et al., 2012). Accordingly, the face was tracked using an extended boosted cascade classifier (Viola & Jones, 2001) implemented in OpenCV library (v. 2.2). Then three blocks of facial components were extracted from the detected face region; left-eye, right-eye and mouth. In order to have the same size of blocks in each image, the detected objects were resized to fixed sizes. The LBPTOP operator was applied on each block separately and the final feature set was created by concatenating the results from each block. Totally 2304 (3 blocks × 3 orthogonal planes × 2^8 local binary patterns) features were extracted from each video segment.

4.2. Extracting HR features from video

The first step was to detect and track the face in the recorded or live video. An extended boosted cascade classifier implemented in OpenCV library (v. 2.2) was used for face tracking. The algorithm focused on the skin regions which are more likely containing PPG signal. Afterwards, the ROI was divided into the RGB channels and the average of each color (RGB) amplitude values is calculated across all pixels in the ROI. These three raw signals are considered as the inputs for the Independent Component Analysis (ICA). Before applying ICA, the raw traces were deterended (Tarvainen, Ranta-aho, & Karjalainen, 2002) and normalized to improve the quality of the signals.

ICA (Comon, 1994) is a special case of Blind Source Separation (BSS) techniques which tries to separate a multivariate signal into statistically independent subcomponents by assuming that the subcomponents are non-Gaussian signals. Here, we adopted a linear ICA based on the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm (Cardoso, 1999). In the linear ICA, it is assume that the observed signals contain linear mixtures of source signals. Typically, ICA cannot identify the actual number of source signals but the number of recoverable sources is less than or equal to the number of observations.

In order to identify the component that contains the HR signals, further analysis was needed. Poh et al. (2010) selected the second component manually as they argued that the HR signal could be observed clearly from that component. Monkaresi et al. (in press) proposed a machine learning method to estimate the HR from the three components automatically. We have used their method to extract HR features. Accordingly, nine features were extracted from the of the three PSD curves of the independent components. The set of features includes: the frequency of highest peaks in the PSD curves before and after applying noise reduction method and the depth of searches in the noise reduction method for each component formed the rest of the features. Then a k-nearest neighbour (k=1) model was trained using the nine features as input vectors and actual HR extracted from recorded ECG. The training and testing process has been done based on a k-fold cross-validation approach (k=10) for each participant. The estimated HR values were used to extract HR features for each video segment. Seven statistical features (mean, median, standard deviation, max, min, range, difference) were extracted from the heart rate estimations for each video segment.

4.3. Classification

Before classification, a correlation-based feature selection (CFS) technique was applied on the extracted features to remove unnecessary features. The CFS evaluates different possible subsets of features and rank them based on its measure (M. A. Hall, 2000). This technique can select a subset of features which are highly correlated with the target parameter and have the lowest internal correlation between features.

In this study, two different channels (Facial expression and physiological signal) were considered and utilizing just a single classifier may not produce optimum result. Previous studies (Hussain, Monkaresi, & Calvo, 2012) also suggested that combining classifiers can increase the accuracy of affect detection in multimodal systems. The *vote* classifier with the average probability rule was used for combining base classifiers. For the base classifiers, we have selected the common types of classifiers which are widely used and showed reasonable performance in affect classification (Nguyen, Bass, Li, & Sethi, 2005). Support Vector machine (SVM), k-Nearest Neighbor (kNN) and Decision Trees are considered as base classifiers.

To evaluate the performance of the classifier, Cohen's Kappa was calculated and reported (Cohen, 1960). Compare to other simple agreement measures (e.g. percent), the Cohen's Kappa is a robust measure because it is less sensitive to the agreement occurring by chance.

5. Results

For each video segment 2304 features were extracted by the LBPTOP method. Heart rate was also extracted using the method proposed in the Section 3.2. Each video segment last 10 seconds and for each second there is an estimation for heart rate. Altogether 2311 (2304 LBPTOP + 7 HR) features were extracted and then synchronized with corresponding annotations (normative ratings and concurrent self-reports). The following section reports the classification accuracies for detecting valence and arousal. The performance of each modality for discriminating between two levels of valence and arousal was also explored.

5.1. Feature analysis

In order to evaluate the importance of each specific feature we have calculated the Chi-squared values with respect to self-reports and normative labels. This analysis showed the relation between different types of features from each channel (HR, and LBPTOP) and participants' affects. Finding the most important features can be useful to build the general model for affect detection. Table 1 shows the Chi-squared values for top 5 features in valence and arousal analysis. According to this table the majority of the features came from the LBPTOP features (60%). However in some cases HR features were contributed more among the top 5 features selected for affect detection. The most surprising result was observed for normative arousal labels. All the top 5 features were from the HR features which indicated the importance of HR features for normative arousal detection. The *HR-median* feature was also the best indicator for normative valence detection. The mean and median values for the measured heart rates also appeared in the top 5 features for self-reported arousal detection.

According to the Chi-squared values, the top five fea-

Table 1. Chi-squared values for top 5 features in affect (valence/arousal) analysis

	Self-reported Valence		Self-reported Arousal		Normative Valence		Normative Arousal	
#	**Chi-Sq.**	**Feat. Name**	**Chi-Sq.**	**Feat. Name**	**Chi-Sq.**	**Feat. Name**	**Chi-Sq.**	**Feat. Name**
1	26.91	P984_R-Eye_XY	32.22	P1973_Mouth_XT	41.68	***HR-median***	47.78	***HR-mean***
2	24.33	P795_R-Eye_XY	23.41	***HR-median***	27.33	P1658_Mouth_XY	44.89	***HR-max***
3	22.17	P820_R-Eye_XY	21.76	P1955_Mouth_XT	24.33	P1650_Mouth_XY	41.63	***HR-median***
4	21.55	P230_L-Eye_XY	21.36	P2031_Mouth_XT	24.31	P2234_Mouth_YT	39.73	***HR-min***
5	21.54	P1684_Mouth_XY	21.08	***HR-mean***	23.73	P1777_Mouth_XY	25.62	***HR-range***

tures related to self-reported valence were all from XY planes. In addition, three out of top five features in normative valence analysis were also from XY planes. These findings indicated that the appearance of the face were more informative for valence detection compared to motion-related features extracted from XT and YT planes. On the other hand, the features extracted from XT planes contributed more in the top five features selected for self-reported arousal detection.

5.2. Classification results

The results are reported in three sub-sections: user-dependent, gender-specific and general (user-independent) models. For each model, the classification tasks were performed for detecting four types of affect representations. Valence and arousal were considered independently with respect to type of labels: self-reported ratings (*selfVal*, *selfAro)* and normative ratings (*normVal*, *normAro)*.

User-dependent analysis

In user-dependant analysis, 17 specific models were trained and tested for each participant. The first row in Table 2 presents the average kappa scores for classifying affects using the HR features and LBPTOP features separately and the fusion model.

For all normative ratings, fusion models achieved the best results with a reasonable accuracy compared to each individual channel. The best result (kappa=0.64) was achieved by the fusion model for classifying two levels of normative valence which was an excellent performance. The improvements of fusion model over the LBP channel were statistically significant in all 2 cases as specified by paired samples t-tests ($p<0.05$). The HR channel was not be able to detect self-reported affects. However the LBP channel obtained a fair accuracy for classifying self-reported affects.

Gender-specific analysis

For this analysis, we separated our dataset into two portions, with one part containing only male participants' data (n=10) and the other part only female participants' data (n=7). Then, data from individual participant were standardized (converted to z-scores) to address individual variations of head behaviour and physiological differences. We built and trained specific models for each of the datasets in order to compare the performance of gender-specific models. A leave-one-participant-out cross-validation approach was used to evaluate these models. The kappa measures for classifying valence and arousal for females and males models are also reported in Table 2 (second and third rows respectively).

For the female participants, the HR channel did not improve the fusion model for detecting self-reported affects. However the HR channel showed better performance for detecting normative affects for female participants. The HR channel improved the kappa measure for detecting normative arousal. A kappa measure of 0.32 was achieved by the fusion model for detecting normative arousal.

According to the third row in Table 2, the HR channel and the fusion model were not successful in classifying self-reported affects for the male participants. But for the normative ratings the fusion model achieved the best results for the male participant model. The major improvement of 0.08 was achieved by adding the HR channel to the LBPTOP features for detecting normative arousal. Among gender-specific models, female model obtained better results for detecting normative affects in particular normative arousal.

User-independent analysis

To build a user-independent model, data from individual participants were first standardized and then combined to yield one large data set with 1020 instances. The kappa measures were calculated based on leave-one-participant-out cross validation approach. The fourth row in Table 2 shows the kappa measures for detecting the affects using the general model.

As expected for the user-independent model, the achieved kappa measures were less than values that obtained using the user-dependent models. Again, the HR channel did not perform well in detecting self-reported affects. On the other hand, the HR channel improved the accuracy of normative valence and arousal detection by 0.02 and 0.09 respectively. The HR channel also obtained positive kappa measures for detecting normative valence and arousal. Combining LBPTOP (kappa=0.08) and HR (kappa=0.06) features showed a supper additive effect for detecting normative arousal (kappa=0.17).

Table 2. The average kappa scores for classifying affects

	Self-reported Valence			Self-reported Arousal			Normative Valence			Normative Arousal		
	HR	**LBP**	**Fusion**	**HR**	**LBP**	**Fusion**	**HR**	**LBP**	**Fusion**	**HR**	**LBP**	**Fusion**
User-dependent	-0.05	0.56	0.55	0.01	0.44	0.44	-0.09	0.61	0.64	-0.12	0.51	0.52
Female-specific	0.00	0.32	0.32	0.02	0.25	0.25	0.06	0.28	0.22	0.15	0.31	0.32
Male-specific	-0.03	0.12	0.12	0.00	0.05	0.02	0.04	0.20	0.20	0.08	0.11	0.19
General model	-0.02	0.13	0.13	-0.04	0.07	0.07	0.08	0.13	0.15	0.06	0.08	0.17

6. Conclusion

This paper introduced a new fusion model for affect detection that is able to extract HR signals from facial videos. The results showed that combining these HR features with other facial expression features (e.g., LBPTOP features) can improve the accuracy of affect detection using the normative rating. The HR channel was more successful in detecting normative arousal for female participants (kappa measure =0.15). The fusion model showed a reasonable accuracy for detecting affect (normative rating) in user-independent analysis as well. This study demonstrates the feasibility of using contact-less physiological signal measurements for affect detection even though the improvement was not too much. Replacement of traditional physiological sensors with a camera can increase the usability of an affect detection system. The approach in this paper is the first attempt in using remote physiological measurement with other modalities for affect detection and more improvements needs to be done. Extracting more physiological signals such as inter-beats intervals (IBI) and respiration rates using video-based method and adding them to the fusion model can be considered for future work.

References

Bailenson, J., Pontikakis, E., Mauss, I., Gross, J., Jabon, M., Hutcherson, C., … John, O. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, *66*(5), 303–317. doi:10.1016/j.ijhcs.2007.10.011

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., … Narayanan, S. (2004). Analysis of Emotion Recognition using Facial Expressions , Speech and Multimodal Information. In *Proceedings of the 6th International Conference on Multimodal Interfaces* (pp. 205–211).

Calvo, R. A., & D'Mello, S. (2010). Affect Detection : An Interdisciplinary Review of Models , Methods , and Their Applications. *Affective Computing, IEEE Transaction on*, *1*(1), 18–37.

Cardoso, J. F. (1999). High-order contrasts for independent component analysis. *Neural Computation*, *11*(1), 157–92.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*(3), 287–314.

D'Mello, S., & Kory, J. (2012). Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In *14th ACM International Conference on Multimodal Interaction* (pp. 31–38). Santa Monica, California, USA.

Fei, J., & Pavlidis, I. (2010). Thermistor at a distance: unobtrusive measurement of breathing. *IEEE Transactions on Bio-Medical Engineering*, *57*(4), 988–98. doi:10.1109/TBME.2009.2032415

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, *30*(4), 1334–1345.

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, *85*(1), 6–23.

Hall, M. A. (2000). Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 359–366).

Hertzman, A. B., & Spealman, C. R. (1937). Observations on the finger volume pulse recorded photo- electrically. *American Journal of Physiology*, *119*, 334–335.

Hussain, M. S., Calvo, R. A., & Pour, P. A. (2011). Hybrid Fusion Approach for Detecting Affects from Multichannel Physiology. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 568–577). Springer.

Hussain, M. S., Monkaresi, H., & Calvo, R. A. (2012). Combining Classifiers in Multimodal Affect Detection. In *Proceedings of the Tenth Australasian Data Mining Conference (AusDM 2012)* (pp. 103–108). Sydney, Australia.

Kim, J., & Andre, E. (2006). Emotion recognition using physiological and speech signal in short-term observation. In E. André, L. Dybkjær, W. Minker, H. Neumann, & M. Weber (Eds.), *Perception and Interactive Technologies. LNCS* (Vol. 4021, pp. 53–64). Springer-Berlin Heidelberg.

Lang, P., & Bradley, M. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *Psychology*.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (p. 61). Gainesville, FL.

Li, C., Cummings, J., & Lam, J. (2009). Radar remote monitoring of vital signs. *Microwave Magazine,*, *10*(February), 47–56.

Monkaresi, H., Calvo, R. A., & Yan, H. (in press). A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam. *IEEE Journal of Biomedical and Health Informatics*, *PP*(99), 1–1. doi:10.1109/JBHI.2013.2291900

Monkaresi, H., Hussain, M. S., & Calvo, R. A. (2012). A Dynamic Approach for Detecting Naturalistic Affective States from Facial Videos during HCI. In M. Thielscher & D. Zhang (Eds.), *AI 2012: Advances in Artificial Intelligence* (pp. 170–181). Berlin, Heidelberg: Springer Berlin Heidelberg.

Nguyen, T., Bass, I., Li, M., & Sethi, I. (2005). Investigation of combining SVM and decision tree for emotion classification. In *The Seventh IEEE International Symposium on Multimedia (ISM'05)* (pp. 540–544).

Nicolaou, M. A., Gunes, H., & Pantic, M. (2011). Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence–Arousal Space. *Affective Computing, IEEE Transactions on*, *2*(2), 92–105.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, *29*(1), 51–59.

Poh, M. Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, *18*(10), 10762–74.

Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal Emotion Recognition in Response to Videos. *IEEE Transactions on Affective Computing*, *3*(2), 211–223. doi:10.1109/T-AFFC.2011.37

Tarvainen, M. P., Ranta-aho, P. O., & Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. *IEEE Trans. Biomed. Eng.*, *49*(2), 172–175.

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics Express*, *16*(26), 21434–45.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '01)* (pp. 511–518).

Zhao, G., & Pietikäinen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Analysis and Machine*, *29*(6), 915–928.