VOL. 11, NO. 1, JANUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

EMOTION DETECTION FROM THERMAL FACIAL IMPRINT BASED ON GLCM FEATURES

Latif M. H, Md. Yusof H., Sidek S. N., Rusli N. and Sado Fatai
Department of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia
E-Mail: myhazlina@iium.edu.my

ABSTRACT

Social intelligence in robots has been demonstrated and recognized in numerous contemporary studies especially for Human Robot Interaction (HRI). However, it has become increasingly apparent that social and interactive skills are prerequisites in any application areas and contexts where robots need to interact and collaborate with other robots or humans. The main focus now shifted on how the robots should perceive human affective states and manifest it through action. Recognition of human affective states could be achieved through affective computing by using numerous modalities such as speech, facial expression, body language, physiological signals etc. There are two approaches to access the affective states; invasive and noninvasive. Decades of researches and findings were mostly focussed on the invasive approach; Electroencephalogram (EEG), heart rate, blood flow, Galvanic Skin Response (GSR) etc. When it comes to affect recognition using noninvasive approach, very few numbers of publications have been done to date. In this paper, we presented an efficient method for thermal image feature extraction using the Gray Level Co-occurrence Matrix (GLCM) technique. By analysing the heat pattern on the facial skin, this work attempts to investigate the suitability of the thermal imaging technique for affect detection. The findings of this study indicate thermal imaging as a contactless and noninvasive method for appraising human emotional states.

Keywords: affect detection, human robot interaction (HRI), thermal image, gray level co-occurrence matrix (GLCM).

1. INTRODUCTION

Emotions are a common term used daily. Although there is various definitions of "emotion", the commonly used term is based on the psychology term where, emotion is defined as "a complex state of feeling that results in physical and behavior". There are two main categories on theory of emotion; cognition and somatic. The cognition category appeal to a necessary element of emotion and the subjective manifestation which could be "conscious or unconscious, intentional or unintentional and take a form of a judgment or a thought" [1]. The somatic category on the other hand relies on somatic features and "seeks to describe emotional expressions and perceptions of emotional expressions [2]. This theory believes that body responses trigger emotional reactions. Most of the researchers only use two-dimension since three-dimension is difficult to apply consistently. According to Russell and Steiger, this approach is "straightforward, simple and provide interval data which can be used in the statistical processing" [3].

Picards [4] states that "Emotion plays an essential role in rational decision-making, perception, learning and a variety of functions". The implicit communication between a communicator plays a significant role in human social interactions. However, in several cases, meaningful communication especially for individual with special attention (e.g. autistic, ADHD, etc.) becomes a nontrivial task. Typical favourable modalities to interaction such as verbal communication or body language may be deterred which cause normal communication; could not be achieved [5]. It would be immensely useful to have a robotic system that is capable of such communication with the operator and modify its behaviour when required. A machine that is capable of recognizing emotional states and synthesize proper response would be very useful for Human Robot Interaction (HRI). Recent studies in artificial intelligence and robotics have proven the ability of social intelligence in robots. Over the years, the trend in emotionally and socially aware system research has been increasing which demonstrates the importance of HRI. Recognition of human affective states could be achieved through affective computing by using numerous modalities such as speech, facial expression, body language, physiological signals etc. Although a lot of information had been gathered and documented, the modes of experimentation are mostly invasive in nature where direct contact between the subject and the sensor is required. This would be intimidating to the subject and restricted their mobility, which somehow would provide a percentage of uncertain information due to stress experienced by the subjects. The maturity and evolution of sensing technique have made contact-less and non-invasive ANS responses monitoring possible through the application of thermal imaging which has shown potential in characterizing the subdivision of ANS. Merla *et al.* [6] have proposed the facial skin temperature exploitability for affective states recognition using thermal imaging.

Countless number of researches has been done in the area of affect recognition using physiological signals. In addition, numbers of excellent review papers on this area also have been published over the years (e.g. [7-8]). Nevertheless, when it comes to affect recognition using thermal signature, very few numbers of publications have been done to date. Table-1 below shows the details of

some researches that have been done so far in the area of affect recognition using thermal signature.

The remaining part of this paper is organized as follows: Section 2 presents the Experimental Design (Stimuli, Induced Emotion, Data Collection and Facial Thermal Imprint). Section 3 describes some processes require for in the affect detection from thermal image (Thermal Image Acquisition, Pre-processing and ROIs Generation, Feature Extraction and Classification). Lastly, section 4 describes on the finding and discussion.

Table-1. Review on affective recognition- thermal imaging approach.

Title/Ref. No.	Input signal	Emotions	Observed areas (ROI)	Stimuli used	Extracted data	Classification
Thermal signatures of emotional arousal: a functional infrared imaging study [6]	Heart rate Breathing rate Galvanic skin response Facial skin temperature Arterial pressure	Stress Fear Pleasure	Periorbital Forehead Lips Wrist palm	Sub-painful electric stimulation of median nerve (wrist) Stroop test Movies	Galvanic skin response Heart beat rate Facial blood perfusion rate	Data observation
Thermal imaging as a way to classify cognitive workload [9]	Facial skin temperature	Alertness Anxiety Fear	Periorbital Left cheek Right cheek Nose Chin Forehead	Cognitive stress test (CST)	Mean reaction time (Cognitive Stress Test) Average temperature (ROI)	Artificial Neural Network (ANN)
Thermal imaging for anxiety detection [10]	Facial skin temperature Respiration	Anxiety Alertness Fearfulness	Periorbital Left cheek Right cheek Nasal Chin Neck	Sit in the dark Auditory Sit in dimly lit room Chewing gum Walking	Mapped RGB Triplet to single value (0-255) (inverse linear rainbow transformation) Pixel averaging (0-255) & standard deviation	Data observation
Classifying affective states using thermal infrared imaging of human face [5]	Facial skin temperature Blood volume pulse (BVP) Respiration	Arousal Valence	Periorbital Supraorbital (Left & Right) Nasal	International Affective Picture System(IAPS)	50% of hottest pixels ROI(mean, variance, skewness, kurtosis and entropy)-time series data Mean temperature time-series (denoised-wavelet-based approach-Daubecies wavelet 4dB)	Fisher Linear Discriminant Analysis (LDA) and Genetic Algorithm (GA)
Facial skin surface temperature changes during a "concealed information" test [11]	Facial skin temperature	Stress Fear	Periorbital	Concealed Information Test (CIT) Zone Comparison Polygraph Test (ZCT)	Max & Min temperature of each pixels between frames (peak hold & valley-hold algorithm) Average pixel intensity of 2 bilaterally symmetric ROI	Binary Logistic Regression
Forehead thermal signature extraction in lie detection [12]	Facial skin temperature	Stress Fear	Forehead	Mock crime scenario based questions	Overlap ratio- estimated by accuracy (ratio of correctly classified pixel) Closeness-Hausdorff distance (mean distance)	Binary classification
Interacting with human physiology [13]	Facial skin temperature Metabolic rate (Energy expenditure) Blood volume	Stress Frustration	Periorbital Forehead	Stroop Color Word Conflict Test	10% of hottest pixels(ROI) Mean temperature-temperature signal (10% of hottest pixels) Compute blood flow (bio- heat model)	Correlation between (Energy expenditure) and thermal data using Pearson correlation factor

2. EXPERIMENTAL DESIGN

The setup for the experiment is illustrated in Figure-1. The distance between subject and the thermal camera was maintain at 1 meter and the emissivity constant parameter for the thermal camera was set to 0.98 (Human skin). During data collection, subject was requested to remove their glasses if they wore any. Subjects were advised not to take any meal 1-hour before data collection to avoid metabolic effect of digestion.

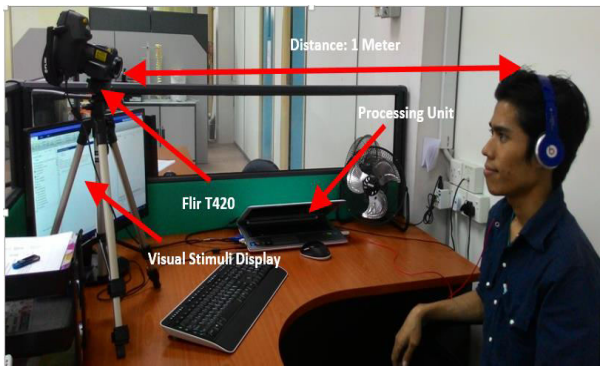

Figure-1. Experimental setup.

A. Stimuli

For the purpose of emotion elicitation, a specific video clips had been administered to the subjects through the monitor. The video clips were chosen because of the sound and music contribution to emotion elicitation as being reported in several studies. Common methods for stimulating affective states include affective picture viewing [5], movie clips [6], startle auditory [14], Stroop Color Word Conflict Test [13], and Mock crime scenario based questions [12,15] etc.

B. Induced emotions

In our experiment, three basic human emotions have been tested named as Happy, Sad and Fear. The selection of these emotions was based on the fact that it was completely opposite to one another.

C. Data collection

The duration for each video clips was around 1 minute. Prior to the session commencement, the subjects were advised to minimize the head movement to avoid out of focus images. The session was recorded with Flir T420 with 30 frame per second (fps). For this experiment, we used an uncooled micro bolometer Focal Plane Array (FPA) LWIR (7.5-13 µm) from FLIR (Model T420). This thermal camera has frame rate of 60Hz and thermal sensitivity (N.E.T.D) of < 0.045 °C at 30°C (NETD stand for Noise Equivalent Temperature Difference).

D. Facial thermal imprint

The human body emits electromagnetic radiation which categorized in infrared region (also known as thermal radiation). The temperature of the body is directly correlated to the wavelength of the radiation emitted. Basically, Infrared (IR) spectrum constitutes of active infrared band (Short-Wave Infrared (SWIR)/Near-IR) and passive infrared band (Mid-Wave Infrared (MWIR) and Long-Wave Infrared (LWIR)). The wavelength for MWIR ranges from 3-5 µm while LWIR ranges from 7-14 µm. For the research that dealing with human body, passive band IR is the best choice as the infrared emitted by human falls under passive infrared band, specifically LWIR. Typical image of the face in thermal is illustrated in Figure-2.

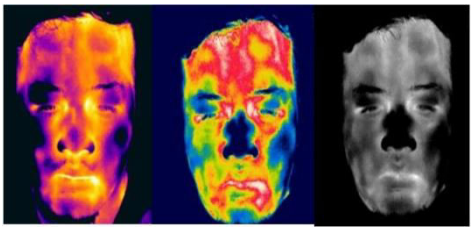

Figure-2. Thermal face

3. PROPOSED FRAMEWORK

The proposed framework adopted in our experiment is shown in Figure-3. The framework consists of three main blocks; Image Processing, Feature Extraction and Test and Validation.

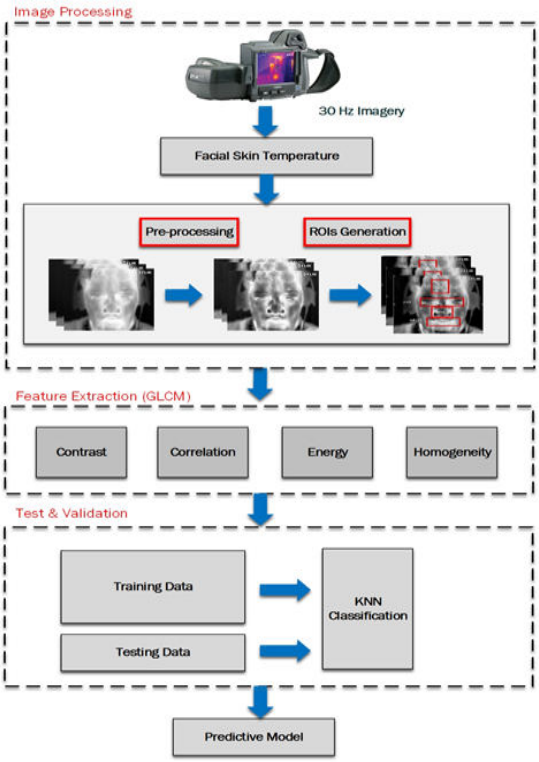

Figure-3. Proposed framework.

3.1 Thermal image acquisition

The acquisition of thermal images was done in a controlled room with relative humidity of 50%, 24 °C and uniform illumination. The subject is seated 1.0 meter from the thermal camera and advice to maintain minimum head movement. The camera's framerate for data collection was set at 30 frames per second (fps). For one set of video, there are 1800 images being captured. The image samples are showed in Figure-4.

Figure-4. Image acquisition.

3.2 Pre-processing and region of interest generation

In the pre-processing stage, the facial thermal imprint images were converted to binary images through thresholding. The process converts the grayscale image I to a binary image. The output image BW replaces all pixels in the input image with luminance lesser than threshold with the value 1 (white) and replaces all other pixels with the value 0 (black). Figure-5 showed the binary image.

Figure-5. Binary image.

The process then continues with blob analysis of the image. In blob analysis, the connected object will have the same label and group accordingly. The size of the area is arranged in descending order and the biggest area which represented the face area was selected. From this process the face area was segmented from the rest of the image. The result from this process is showed in Figure-6.

Figure-6. Segmented face.

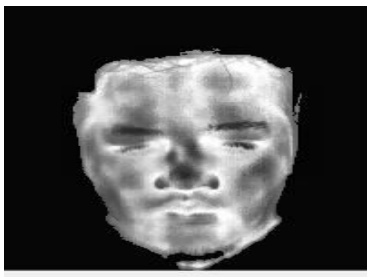

Figure-7. Enhanced face.

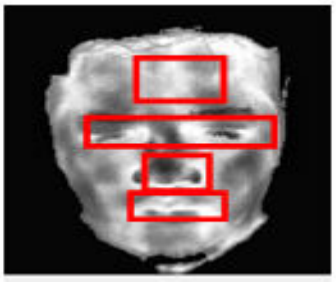

Figure-8. Segmented ROI.

Next, Figure-7 showed the enhanced facial image after normalized using contrast limited adaptive histogram equalization (CLAHE) [16]. The algorithm applies histogram equalization to each sub-region in the image as illustrated by Equation (1). In the equation, M and N are the numbers of pixels and gray level bins in each sub-region. The histogram for each sub-region is represented as h. In order to improve the contrast of the image without increasing the noise, clip limit threshold is set (0.01 in our test), CLAHE redistribute each histogram so that it maintain below the threshold level. In case of gray level exceeding the clip limit and they are distributed uniformly below the threshold level. Lastly, each sub-region is combining through bilinear interpolation. The region of interests (ROIs) were shown in Figure-8; Supraorbital, Periorbital, Nose and Mouth. It was very unfortunate that the automated algorithm for facial parts detection such as Viola-Jones could not be used (we tried). The problem arose in the implementation of Viola-Jones algorithm was multiples ROIs detected on the same region. Changing the merge threshold values did not improve the detection. All of these ROIs were segmented manually by specifying the coordinate of the bounding box around each ROIs. From these ROIs, the GLCM will later be extracted as discussed in next sub-section.

$$I(b) = \frac{(N-1)}{M} \; x \; \sum_{k=0}^{b} h(k) \quad (1)$$

3.3 Feature extraction

For the feature extraction, a widely used texture based feature extraction known as Gray Level Co-occurrence Matrix (GLCM) has been chosen. The GLCM characterize second order statistic of an image by computing how often pairs of pixel with specific values

and in a specified spatial relationship occur in an image. The features are generated by calculating the features for each one of the co-occurrence matrices obtained by using the directions 0°, 45°, 90°, and 135° and specific displacement distances. For the computation of GLCM, we used four displacement distances, d = 1, 2, 3 and 4 and four orientation, θ = 0°, 45°, 90°, and 135°. Hence, for each ROI, there are (64 features x 1200 image = 76800 features). The Anger, Disgust and Surprise dataset have 76800 features respectively which leads to a total of 460800 features overall for each ROI. The dataset (EmotionData) then represented in the form of 460800 features x 4 ROIs (Supraorbital, Periorbital, Nasal, and Mouth).

3.4 Classification

In this experiment, we had tested several classification algorithms for the sake of achieving the best classification result. Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Tree just to name a few were among the classifier tested. The classification process was performed by using the Classification Learner in Matlab 2015a. The dataset Emotion Data, 460800 x 4 features of size were fed into the classifier. The ROIs served as the predictor while the Emotion Lable served as the response of the classifier. Besides, the 10-fold cross validation was chosen for the model validation. There are three options available for model validation in Classification Learner; k-fold cross validation, holdout and no validation. The dataset were divided equally into 10 partitions and the cumulative errors from each fitting (10-folds) were computed to get the k-loss error for the validation stage. The small value of k-loss error could directly reflects the performance of the classifier.

4. RESULT AND DISCUSSIONS

As mentioned in previous section, in this experiment we tested several classification technique to get the best classification result for the three basic emotions; Anger, Disgust and Surprise. Table-2 shows the comparison for the classifier performance.

Table-2. Classifier performance comparison.

Classification method	Feature	Accuracy (%)
SVM: Fine Gaussian	GLCM	88.5
SVM: Medium Gaussian	GLCM	82.3
Tree Complex	GLCM	61.4
Tree: Medium	GLCM	60.2
KNN: Fine KNN	GLCM	88.5
KNN: Weighted KNN	GLCM	99.1

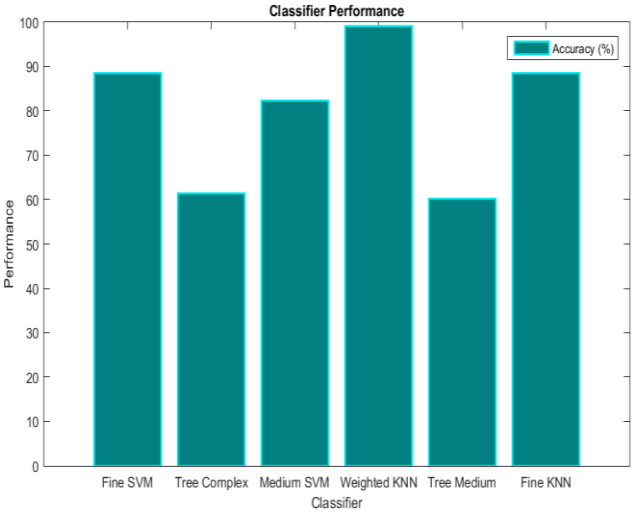

Figure-9. Classifier performance.

For Table-2, the highest classification accuracy was showed by Weighted KNN with 99.1% overall accuracy and followed by Fine KNN with 88.5% overall accuracy. For the Weighted KNN, the numbers of neighbors used was k=678 which based on the rule of thumb for optimal k-value selection. The rule stated that the value of k should be approximately equal to the square root of the instances available in the dataset. The others parameters set were Euclidean as the distance metric) and Squared Inverse as the distance weight. Figure-10 elucidate the confusion matrix for the Weighted KNN. From the confusion matrix, important parameters such as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), True Positive Rate (TPV), False Negative Rate (FNR), Positive Predictive Rate (PPR), and False Discovery Rate (FDR) could be extracted for performance evaluation.

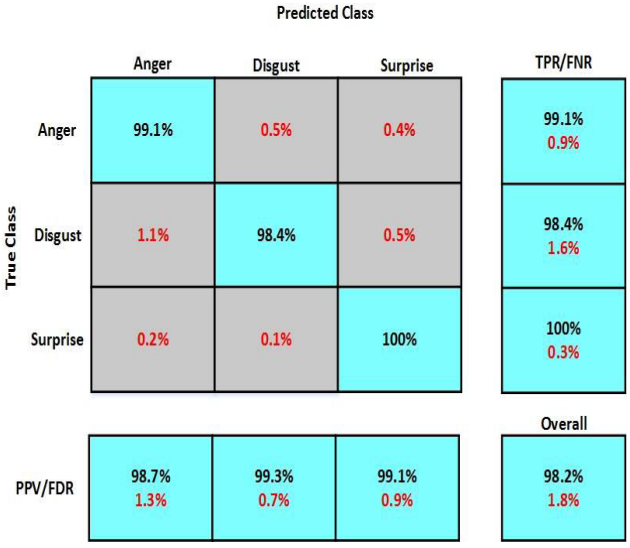

Figure-10. KNN weighted: Confusion matrix.

5. CONCLUSIONS

This research is far from being completed. The initial findings are very encouraging and we are positive that thermal imaging could provide contactless, non-invasive alternative for affect detection in Human Robot Interaction (HRI). However, several known limitations of thermal imaging for real life implementation should be taken care of such as the occlusion of hair-bang and spectacle, the metabolic effect of digestion, drugs, fever, etc. In our experiment and other researches that have been done in the domain of affect detection using non-invasive approach (Thermal Imaging, ECG, GSR etc.), all the data were collected in a controlled room (temperature, humidity, luminance, surrounding noise etc.)

REFERENCES

[1] Lopatovska and I. Arapakis. 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. Information Processing and Management. vol. 47, pp. 575-592.

[2] Zajonc R.B. 1984. Affect Theory: Approach to emotion. Hilldale, Lawrance Erlbaum Associates. New Jersey. pp. 239-246.

[3] J. A. Russell and J. H. Steiger. 1982. The structure in persons' implicit taxonomy of emotions. Journal of Research in Personality. 16: 447-469.

[4] W. R. Picard. 2003. Affective computing: Challenges. International Journal of Human-Computer Studies - Application of affective computing in human-Computer interaction. 59: 55-64.

[5] B. R. Nhan and T. Chau. 2010. Classifying Affective States Using Thermal Infrared Imaging of the Human Face. IEEE Trans. Biomedical Eng. 57(4): 979-987.

[6] Merla and G.L Romani. 2007. Thermal signature of emotional arousal: a functional infrared imaging study. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite International, Lyon, France.

[7] R. A. Calvo. 2010. Affect detection: an interdisiplinary review of models, methods, and their appplications. IEEE Trans. On Affective Computing.1 (1):19-37.

[8] Jerritta S., M. Murugappan, R. Nagarajan and Khairunizam Wan. 2011. Physiological signals based human emotions recognition: A Review. IEEE 7th International Colloquium on Signal Processing and its Applications, pp. 410-415.

[9] J. Steberger, R.S. Allison and T. Schnell. 2010. Thermal imaging as a way to classify cognitive workload. In: Canadian Conference on Computer and Robot Vision. pp. 231-238.

[10] Pavlidis, J. Levine, and P. Baukol. 2001. Thermal Image Analysis for Anxiety Detection. In: Proceeding of the 2001 IEEE International Conference on Image Processing. Thessaloniki, Greece. 2: 315-318.

[11] A. Pollina, A. B. Dollins, S. M .Senter, T. E. Brown, I. Pavlidis, J. A. Levine and A. H. Ryan. 2006. Facial skin surface temperature during a "concealed information" test. Annals of Biomedical Eng. 34(7): 1182-1189.

[12] Z. Zhu, P. Tsiamyrtzis and I.Pavlidis. 2007. Forehead thermal signature extraction in Lie Detection. In: Proceeding of the 29th Annual International Conference of the IEEE EMBS Cite Internationale, Lyon, France. pp. 243-246.

[13] Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey. 2007. Interacting with human physiology. Comput. Vis. Image Understanding. 108(1-2): 150–170.

[14] D. Shastri, A.Merla, P. Tsiamyrtzis, and I. Pavlidis. 2009. Imaging facial signs of neurophysiological responses. IEEE Trans. Biomed. Eng. 56(2): 477-484.

[15] Pavlidis and J. Levine. 2002. Thermal image analysis for polygraph testing. IEEE Engineering in Medicine and Biology Magazine, 21(6): 56-64.

[16] T. Bourlai, A.Ross, C.Chen and L.Hornak. 2012. A Study on Using Mid-Wave Infrared Images for Face Recognition. In Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring and Biometric Technology for Human Identification, Proc. of SPIE. 8371.