Healthcare System Focusing on Emotional Aspect Using Augmented Reality: Emotion Detection by Facial Expression

Somchanok Tivatansakul[a], Gantaphon Chalumporn[b], Supadchaya Puangpontip[b]
Yada Kankanokkul[b], Tiranee Achalakul[b], and Michiko Ohkura[c]
[a]Graduate School of Engineering
Shibaura Institute of Technology, Japan

[b]Department of Computer Engineering
King Mongkut's University of Technology Thonburi, Thailand

[c]College of Engineering
Shibaura Institute of Technology, Japan

ABSTRACT

Current research includes many proposals of systems that provide assistances and services to people in the healthcare fields; however, these systems emphasize the support physical rather than emotional aspects. Emotional health is as important as physical health. Negative emotional health can lead to social or mental health problems. To cope with negative emotional health in daily life, we propose a healthcare system that focuses on emotional aspects. This system provides services to improve user emotion. To improve user emotion, we need to recognize users' current emotional state. Therefore, our system integrates emotion detection to suggest the appropriate service. This system is designed as a web-based system. While users use the system, facial expression and speech are detected and analyzed and to determine the users' emotions. When negative emotions are detected, our system suggests that the users take a break by providing services (designed to provide relaxation, amusement and excitement services) with augmented reality and Kinect to improve their emotional state. This paper focuses on feature extraction and classification of emotion detection by facial expression recognition.

Keywords: E-Healthcare, Emotion Recognition, Facial Expression

INTRODUCTION

In recent years, the design and implementation of ubiquitous, intelligent space and healthcare systems has become very popular in the field of human computer and human robot interaction. Such systems automatically monitor both the environment and the humans within it to provide assistance and services. Several systems provide support of the physical aspects of people at the expense of the emotional aspects. However, emotional aspects are equally important, as negative emotional health can lead to social and mental health issues.

To cope with negative emotions in daily life and to improve emotional states, we designed a new healthcare system that focuses on emotional aspects. Our system integrated emotion detection to recognize users' current emotional state to provides services to increase their positive emotions and reduce their negative emotions (Tivatansakul and Ohkura, 2013).

Emotion detection or emotion recognition is essential and useful in human computer and human robot interaction applications because emotions indicate human feelings and needs (Haq and Jackson, 2011). The application can recognize user emotion in order to provide the appropriate services.

Various techniques have been proposed to recognize and classify user emotion. The three main categories of such techniques are:

- *Speech Emotion Recognition*: Identifies emotional states voice analysis. Several feature extraction methods that are useful for emotion recognition have been proposed such as pitch, formants, and short-term energy (Ververidis and Dotropoulos, 2006). However, the main limitation of this approach is that speech is necessary to recognize user emotion. Therefore, if users do not speak, his/her emotion cannot be determined.
- *Emotion Recognition from Facial Expression*: Recognizes human emotion from facial muscle movement, the movement, eyes, mouths or eyebrows movement and facial texture. Several studies have applied computer vision systems to automatically analyze and recognize changes in facial motion from visual information (Ayyaz, 2012). However, this approach is limited because users are required to animate their emotions with their face and a camera must capture the users' frontal view in order to detect and recognize user emotions.
- *Emotion Recognition using Biological Signals*: Analyzes and recognizes human emotional state from biological signals such as electroencephalography (EEG), electrocardiography (ECG), temperature, and galvanic skin response (GSR). This approach can recognize users' emotion when the users are silent or do not show their emotion on their face or outside appearance. However, this approach is also limited because users must be attached to biological sensors (Rattanyu and Mizukawa, 2011).

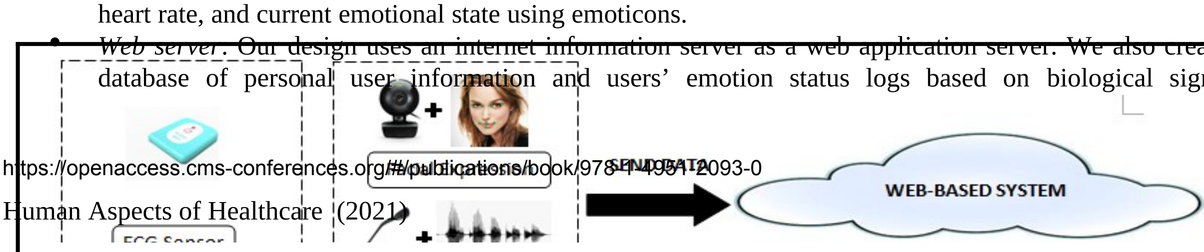
Using facial expression and speech to detect emotion is more suitable for our healthcare system because these approaches recognize the emotions from a natural user interface (face and voice) and thus, biological sensing equipment is not required. Moreover, these approaches are more suitable for use in a work space. This paper focuses on the feature extraction and classification of emotion detection by facial expression recognition.

THE EMOTIONAL HEALTHCARE SYSTEM

We provide a summary of our proposed system in this section (Tivatansakul and Ohkura, 2013). We designed our scheme as a web-based system that users can easily access by the web browsers on a personal computer, a tablet, or a smartphone (Figure 1). While using our system, a webcam and microphone are used to detect users' face and voice in order to recognize and classify their emotions. When these emotions are negative, our system suggests that the users take a break by providing services with augmented reality that are designed to improve their emotional state. After selecting a service, the users can choose the application they wish. To process each application, the users must show an augment reality (AR) marker to the camera to display a virtual object whose purpose is to encourage such positive emotions as relaxation, amusement, and excitement. For some applications, interaction with virtual objects may help decrease negative emotions.

The design framework is shown in Figure 2. This healthcare system consists of five parts:

- *I/O (input/output) devices*: ECG, personal computer, webcam, microphone, Kinect, tablet, speaker, and smartphone.
- *Detection module*: Three applications are designed to detect and analyze users' emotional state, respiration, and gestures with the webcam and microphone, ECG sensor, and Kinect, respectively.
- *Application module*: The application module currently consists of four applications: The augmented reality application is for detecting the AR Marker and displaying virtual models; the breathing control application applies deep breathing techniques to decrease user stress; the notification application suggests that users take a break by offering a service; and the report application displays users' biological information, such as heart rate, and current emotional state using emoticons.
- *Web server*: Our design uses an internet information server as a web application server. We also create a database of personal user information and users' emotion status logs based on biological signals.



- *Emotional Services*: Based on expected user emotions, our system provides three services: relaxation, amusement, and excitement services

EMOTION DETECTION BY FACIAL EXPRESSION

Facial expression is one of the most powerful, natural and easiest methods for human to express and communicate emotion and intention (Shan, 2009). Emotion detection by facial expression recognizes and interprets human emotion from facial muscle movement, eyes, mouth or eyebrows movement, and facial texture.

The workflow of the real-time emotion detection by facial expression proceeds as follows (Figure 3).

- The emotion detection detects a user' face from each video frame (input). To detect the face, we applied EmguCv face detection using Haar Cascades (EmguCv, 2013).
- The detection extracts the facial features and normalizes these features to form feature vectors.
- It then classifies user emotion into one of seven classes (neutral, happy, sad, angry, disgust, fear and surprise) using a classifier that is generated from trained data.
- Finally, it calculates the percentage of each emotion and displays every two seconds.

Feature extraction

To recognize user emotions from facial expression, computer vision and image processing techniques are applied to extract facial features. Feature extraction is an important process of extracting facial features that can represent changes in facial expression caused by emotions. There are two mainly categories for facial features extraction (Tian, 2011):

Geometric-based approaches

Geometric-based approaches represent the geometry of the human face by extracting shapes, points and location of facial components (eyes, eyebrows, mouth and nose) and computing the distances between them to form a feature vector (Zhang, 1998). Geometric-based approaches require accurate and reliable facial component detection which is difficult to obtain with high accuracy with a real-time system that is used in various situation and environments (Jabid, 2010).

Appearance-base approaches

Appearance-based approaches represent facial textures by extracting the change in face appearance and skin texture. Examples of appearance-based approaches are as follows:

- *Local Binary Patterns (LBP)* (Shan, 2009)
  LBP is famous texture description method. The LBP operator encodes the information of curves, edges, spots, and other local features as binary numbers by comparing a 3 x 3 neighborhood pixel with the center value. The LBP operator then divides the image into 7 x 6 rectangular regions to construct a 256-level histogram of each region. All resulting histograms are concatenated to build a global description of the entire face as a feature vector. The LBP operator is shown in Figure 4a.
- *Local Directional Patterns (LDP)* (Jabid, 2010)
  Instead of encoding the intensity of each image pixel, the LDP operator applies an edge detector to compute the edge response in eight directions and encode the texture information by considering the edge responses. Moreover, in case of noise or non-monotonic illumination changes, LDP is more robust than LBP because LDP uses edge responses that are more stable than intensity values in order to generate the binary patterns. The LDP operator is shown in Figure 4b.
- *Directional Ternary Patterns (DTP)* (Kabir, 2013)
  DTP operator also applies an edge detector to compute edge responses in eight directions similar to the LDP. However, the LDP operator encodes the local texture by assigning a 2-value code (0 or 1). The DTP operator, on the other hand, assigns a 3-value code (-1, 0, 1) by differentiating between the smooth and high edge responses using threshold to form positive and negative binary patterns. The DTP operator is shown in Figure 4c.

Our approach

Our feature extraction improves the Directional Ternary Pattern (DTP) which is appearance-based approach. The DTP operator divides the image into 7 x 6 rectangular regions and then generates positive and negative binary patterns by comparing edge responses with the threshold. The edge responses are generated from the edge detection process using Robinson eight-directional masks. The DTP operator then concatenates 256-level histograms of positive and negative binary patterns to construct feature vectors with 512 lengths for each region. The DTP feature vector is larger than LBP or LDP feature vectors, so the DTP process is slower than these processes for classification. To improve the DTP performance and accuracy, our approach applies a sign bit to calculate the ones' complement of binary numbers in order to generate positive and negative binary patterns, as shown in Figure 4d. We then find the 128-level histogram of positive and negative binary patterns and concatenated them to construct feature vectors with 256 lengths.

Our approach descriptor generates a binary pattern (BP) by comparing the edge responses with the threshold (t = 40). The threshold is set to 40 based on the standard deviation of edge responses from smooth face regions (Kabir, 2013). The descriptor then splits the BD into a positive binary pattern (PBD) and negative binary pattern (NBD) in order to generate the ones 'complement PBD and NBD. The descriptor is illustrated below.

$$\text{Ones' complement } \mathrm{PBD}_x = \sum_{i=0}^{7} \mathrm{Comp}\left(\mathrm{PBD}_x\ (\mathrm{i})\right) * 2^i, Comp(a) = \begin{cases} a, \mathrm{PBD}_x\ (0) = 0 \\ 0, \mathrm{PBD}_x\ (0) = 1 \text{ and } a = 1 \\ 1, \mathrm{PBD}_x\ (0) = 1 \text{ and } a = 0 \end{cases} \tag{1}$$

$$\mathrm{PBD}_x\ (\mathrm{i}) = \begin{cases} 1, EdgeResponse(i) \geq t\ ; t = 40 \\ 0, EdgeResponse(i) < t\ ; t = 40 \end{cases} \tag{2}$$

$$\text{Ones' complement } \mathrm{NBD}_x = \sum_{i=0}^{7} \mathrm{Comp}\left(\mathrm{NBD}_x\ (\mathrm{i})\right) * 2^i, Comp(a) = \begin{cases} a, \mathrm{PBD}_x\ (0) = 0 \\ 0, \mathrm{PBD}_x\ (0) = 1 \text{ and } a = 1 \\ 1, \mathrm{PBD}_x\ (0) = 1 \text{ and } a = 0 \end{cases} \tag{3}$$

$$\mathrm{NBD}_x\ (\mathrm{i}) = \begin{cases} 1, EdgeResponse(i) < t\ ; t = -40 \\ 0, EdgeResponse(i) \geq t\ ; t = -40 \end{cases} \tag{4}$$

PERFORMANCE EVALUATION

The universal or basic emotions in facial expression consist of seven emotions: neutral, happy, sad, angry, disgust, fear and surprise, as shown in Figure 5. These emotions are the emotions that people of most nationalities express with their face in the same way (Ekman, 1974). Several facial emotion recognition systems recognize universal emotions by measuring the recognition performance of a six-class emotional expression set (happy, sad, angry disgust, fear and surprise) and seven-class emotional expression set that includes a neutral emotion. We will evaluate the performance of our emotion detection by facial expression by comparing the accuracies of the six-class and seven-class emotional expression sets with the other feature extraction methods of LBP, LDP and DTP.

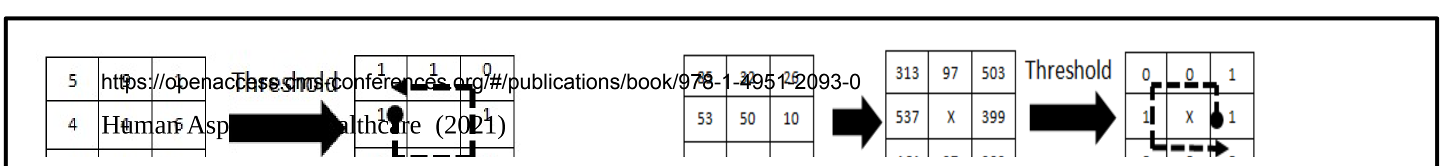


Figure 4. Feature extraction methods.

Figure 5. Seven basic emotions in facial expression from extended CK+ dataset (Lucey, 2010)

Facial expression data

For facial expression data, we used an extended Cohn-Kanade Dataset (CK+) as shown in Figure 5. The CK+ (Lucey, 2010) is an extension of the Cohn-Kanade Dataset that consists of 593 sequences of 123 subjects that are FACS (Facial Action Coding System) coded at the peak frame. All sequences are from the neutral face to the peak expression. However, only 327 of the 593 sequences have emotion sequences because the peak expression of 327 sequences fits the prototypic emotion definition that was validated with reference to the FACS Investigators Guide and confirmed by visual inspection by emotion researchers. Each peak expression of each sequence is assigned to have only one emotion. Therefore, we selected 309 emotion sequences for which each peak expression is labeled as one of the six basic emotions. In each emotion sequence, we selected three peak expressions. We also selected one neutral expression for each subject from the emotion sequences. Therefore, we selected 1033 images in total (207 happy, 177 disgust, 75 fear, 135 angry, 84 sad, 249 surprise and 106 neutral).

Classification Method

Multiclass Support Vector Machine (SVM) classification was applied to recognize users' emotion. SVM was originally proposed by Vladimir N. Vapnik in 1995 (Prasartvit, 2012 and Beniwal, 2012). SVM is a well-known supervised learning model for both linear and nonlinear data classification that has been successfully applied in such computer vision as pattern recognition, image classification and text categorization because of its favorable performance. SVM uses a hyper plane to separate the classes and maximize the margin between the classes. The hyper plane is built from margins and the nearest data points called support vectors which are the critical elements of the training set. The mechanism that defines the mapping process is called kernel function. The kernel function can be linear and non-linear, such as polynomial or Gaussian functions.

RESULTS

For our performance evaluation, we performed ten-fold cross-validation with the SVM classifier using RapidMiner Studio 5 (RapidMiner, Inc., 2013) which is an application for the design of analysis process. The training and testing datasets are generated from the facial feature extraction of the six-class and seven-class emotional expression sets using our approach, LBP, LDP and DTP. The results of the performance evaluation of the emotion detection by facial expression are shown in Tables 1 to 3.

From the results in Table 1, our approach was more accurate than LBP, LDP and DTP for both the six-class and seven-class emotional expression sets using extended CK+ dataset. For the six-class emotional expression classification (Table 2), the recognition rates of each emotion recognized by our approach were: angry (100%), disgust (100%), fear (94.67%), happy (99.52%), sad (92.86%) and surprise (98.80%). For the seven-class emotional expression classification, the accuracy of fear increased and angry, disgust and happy remained the same, but the accuracy of sad and surprise expressions decreased as they were confused with the neutral expression. In conclusion, however, the results from Tables 1 through 3 confirm that our approach improves the accuracy and performance of facial emotion recognition.

DISCUSSION

For emotion recognition by facial expression, speech and biological signals, feature extraction and classification methods are very important to produce higher accuracy and reliable performance. Our emotion detection from facial expression improves the DTP feature extraction method by applying sign bit to calculate the ones' complement of a binary pattern to form feature vectors that represent changes in facial expression caused by emotions. Furthermore, we applied a SVM classifier because it recognizes emotions with high accuracy (Shan, 2009). As a result, our emotion detection by facial expression produced accuracies of 98.49% for six-class emotional expression set and 95.94% for seven-class emotional expression set. Moreover, our approach can recognize the angry and disgust expressions at up to 100 %.

With the six-class emotional expression set (Table 2), the fear expression was confused with the surprise expression because the expression in the eyes and mouth were similar in some test sets. In addition, the sad expression was confused with the angry expression for similar reasons. Therefore, it is difficult to distinguish between sad and angry expressions in this dataset. For seven-class emotional expression set, which includes a neutral expression (Table 3), the recognition accuracy for the sad and surprise expressions decreased because of confusion with the neutral expression. In addition, the neutral expression was often confused with the angry expression because of similarities in facial expression. In summary, the emotion detection by facial expression applies a pattern recognition technique to detect and recognize user emotion. This method may have difficulty classifying similar facial expressions that are caused by different emotions.

We evaluated the performance of our approach with only one dataset in which most of the subjects are Euro-American and the numbers for each emotion expression are not equal. In the future, we plan to evaluate with datasets that consist of mostly Asian subject and similar numbers for each emotion expression in order to avoid bias for some emotions.

Table1: Comparison between our approach, LBP, LDP and DTP using linear SVM.

Feature Extraction Methods	6-Class emotional expression set	7-Class emotional expression set
Local Binary Pattern (LBP)	89.65% +/- 2.70%	84.41% +/- 3.29%
Local Directional Pattern (LDP)	98.49% +/- 1.20%	94.10% +/- 0.51%
Directional Ternary Pattern (DTP)	98.38% +/- 1.30%	94.58% +/- 1.80%
Our Approach	98.49% +/- 0.71%	95.94% +/- 1.42%

Table2: Confusion matrices of 6-class expression recognition of our approach using linear SVM.

	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)
Angry	**100**	0	1.33	0.48	7.14	1.2
Disgust	0	**100**	0	0	0	0
Fear	0	0	**94.67**	0	0	0
Happy	0	0	0	**99.52**	0	0
Sad	0	0	0	0	**92.86**	0
Surprise	0	0	4.00	0	0	**98.80**

Table3: Confusion matrices of 7-class expression recognition of our approach using linear SVM.

	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Sad (%)	Surprise (%)	Neutral (%)
Angry	**100**	0	0	0.48	3.57	0	14.15
Disgust	0	**100**	0	0	0	0	2.83
Fear	0	0	**97.33**	0	1.19	0	0
Happy	0	0	0	**99.52**	0	0	0
Sad	0	0	0	0	**88.10**	0	3.77
Surprise	0	0	2.67	0	0	**98.39**	2.83
Neutral	0	0	0	0	7.14	1.61	**76.42**

CONCLUSIONS

To cope with negative emotional health in daily life, we proposed a new healthcare system that focuses on emotional aspects. This system integrates emotion detection from facial expression and speech in order to understand user emotions. To analyze and detect emotions from facial expression, we improved the DTP approach. Our approach applies a sign bit to calculate the ones' complement of binary number in order to generate positive and negative binary patterns to constructs a 256-level histogram as feature extraction for emotions classification. The results confirm that our approach improves the accuracy and performance of facial emotion recognition.

REFERENCES

Beniwal, S., Arora, J. (2012), "*Classification and feature selection techniques in data mining*," INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 1 No. 6, pp. 1-6.

Ekman, P. (1974), "*Universal facial expressions of emotion,*" Culture and Personality: Contemporary Readings/Chicago, pp. 151-158.

EmguCv Website: http://www.emgu.com/wiki/index.php/Main_Page

Haq, S., Jackson, P. J.B. (2011), "*Multimodal Emotion Recognition,*" Machine Audition: Principles, Algorithms and Systems, pp. 398-423.

Hussain, A., Khan, S. M, Nazir, M., Iqbal, M. A. (2012), "*Survey of various feature extraction and classification techniques for facial expression recognition*," In Proceedings of the 11th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications, and Proceedings of the 11th WSEAS International Conference on Signal Processing, Robotics and Automation, and Proceedings of the 4th WSEAS International Conference on Nanotechnology; World Scientific and Engineering Academy and Society (WSEAS), pp. 138-142.

Jabid, T., Kabir, M., Chae, O. (2010), "*Robust Facial Expression Recognition Based on Local Directional Pattern*," ETRI JOURNAL Volume 32 No.5, pp. 784-794

Kabir, M. H., Ahmed, F. (2013), "*Face recognition with directional ternary pattern (DTP)*," In 2012 International Conference on Graphic and Image Processing.

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I. (2010), "*The Extended Cohn-Kanade Dataset (CK+): A* *complete dataset for action unit and emotion-specified expression*", Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp.94-101.

Prasartvit, T., Kaewkamnerdpong, B., Achalakul, T. (2012), "*Dimensional reduction based on artificial bee colony for classification problem*", In Bio-Inspired Computing and Applications, pp. 168-175.

RapidMiner, Inc. (2013), *RapidMiner Studio 5*. Website: http://rapidminer.com/

Rattanyu, K., Mizukawa, M. (2011), "*Emotion Recognition Based on ECG Signals for Service Robots in the Intelligent Space During Daily Life*," JOURNAL OF ADVANCED COMPUTATIONAL INTELLIGENCE AND INTELLIGENT INFORMATICS Volume 15 No.5, pp. 582-591.

Shan, C., Gong, Sh., McOwan, P. W. (2009), "*Facial expression recognition based on Local Binary Patterns: A comprehensive study*," IMAGE AND VISION COMPUTING Volume 27 No. 6, pp. 803-816.

Tian, Y. L., Kanade, T., Cohen, J. F (2011), "*Chapter 19 Facial Expression Recognition*", Handbook of Face Recognition, Springer-Verlag London Limited, pp. 487-519.

Tivatansakul, S., Ohkura, M. (2013), "*Healthcare system design focusing on emotional aspects using augmented reality — Relaxed service design*," 2013 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), pp.88-93.

Ververidis, D., Kotropoulos, K. C. (2006), "*Emotional speech recognition: Resources, features, and methods*", SPEECH COMMUNICATION Volume 48 No. 9, pp. 1162-1181.

Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S. (1998), "*Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron,*" Third IEEE International Conference on Automatic Face and Gesture Recognition, pp.454-459.