

The Emotion Recognition System with Heart Rate Variability and Facial Image Features

Pei-Yang Hsieh and Chiun-Li Chin
The Department of Applied Information Science
Chung Shan Medical University
Taichung, Taiwan
calvin147258@hotmail.com
ernestli@csmu.edu.tw

Abstract—Amid the rapid advance of technologies in recent years, people live at fast paces and suffer invisible pressures in family, at workplace as well as in school. Emotional undulation tends easily to lead to melancholy. Statistically, the number of bipolar disorder sufferers in the last three years was 40 times that in the past 10 years. What should be more attended to is that 90 percent of suicide cases involved a history of “depression” or “bipolar disorder”. Currently, to assess the mental pressure on an individual, questionnaires or interviews are applied and determination is made by physician’s personal experiences. The development of a more viable quantitative method for assessing individual psychological pressure would facilitate self-management of health. This paper thus aims to CCD-capture human face, measure user’s physiological data using BtECG and determine his/her emotional status by using the fuzzy theory based on the analysis of human face, eyes and Heart Rate Variability. The features of measurements and the determination are also fed to database, which can serve as reference for doctors in consulting related psychological disorders in future.

Keywords-component: *Bipolar Disorder, Fuzzy Inference, Heart Rate Variability, BtECG, Emotion State*

I. INTRODUCTION

Living at fast paces and in a hectic society, people today suffer invisible pressures in family, at workplace as well as in school. They find themselves disoriented from motional undulation and, additionally, are liable to be in an ill cycle of melancholy due to complex interpersonal competitions which arise when mental adaption or external support is not complete. As a result, working efficiency drops, more frustration and setbacks follow with the ill cycle, causing not only psychological problems but also physiological disorders.

The object of this paper is to develop an emotion recognition system by combining the analysis of Heart Rate Variability (HRV) and the recognition of facial expressions. As psychological disorders are relatively difficult to detect in normal conditions, it is hope to determine, as early as possible, the presence of these disorders by measurement of HRV and recognition of facial expressions. Or, otherwise, the abnormal emotions due to illness are hoped to be diagnosed as early as possible, in order to prevent tragedies. With the development of the system, not only is it possible to measure user’s emotional status, but also can the measurements be saved in the database as future reference for doctors in case medical aid is sought.

II. DOCUMENTARY REVIEW AND EXPLORATION

Sufferers of psychological disorders like depression and bipolar disorder increase in recent years. According to statistics, the number of bipolar disorder sufferers in the last three years was 40 times that in the past 10 years. More importantly, 90 percent of suicide cases involved a history of “depression” or “bipolar disorder”. According to data from the Suicide Prevention Center of Department of Health, averagely one person commits suicide in every two hours and two minutes. Last year 1,472,000 Taiwan residents “thought of” ending their lives, with 416,000 people “did attempt to end their lives”, of whom 130,000 “attempted more than once”. Depression occurring on children or teenagers generally differs from that on adults: children tend not to talk about their own symptoms or describe their moods using the terms like melancholy; but they are more likely to express it through behaviors. In contrast, the depression in the elderly is obscure partly because it is traditionally believed that the elderly are losing their mobility and look less lively such that they are often thought to be simply getting old by their offspring, with the potential symptoms of depression being ignored. Hence the development of emotion recognition system becomes necessary.

Reference is first made to Heart Rate Variability, HRV. Heart Rate Variability means the beat-to-beat variation in time within a certain period of sinus rhythm, i.e., the analysis of the magnitude and regularity of the variation in heart rate [1]. Unlike mean heart rate variance which is expressed in minute, the HRV analysis examines the nuance time variance in each cycle of heartbeat and its regularity. Because heart rate is subject to many effects like breathing, blood pressure, endocrine and emotion, it can be used to observe the interaction between autonomic nerve and sino-atrial node. As well as a tool to analyze the conditions of cardiovascular diseases, it is a valuable index in predicting sudden cardiac death and cardiac arrhythmia.

The common methods of HRV analysis include Time Domain Methods and Frequency Domain Methods [1].

1. Time Domain Methods:

This can be divided in statistic and diagram approaches. Overall, time domain methods use indices that depict the shapes of standard deviation of all normal to normal R-R intervals (SDNN), and the root mean square of successive

differences between adjacent normal cycles (RMSSD), quantitatively HRV RMSSD with broader shape indicates bigger HRV, and vice versa.

2. Frequency Domain Methods:

The common methods of transforming HRV curves into spectra include Fast Fourier Transform (FFT). Frequency domain methods make differences in the meaning of analytic results between long periods and short ones. With analyses of shorter periods (5min), it needs to control the well surrounding conditions so as to avoid that affecting autonomic nervous activities. With analyses of longer periods (24hr), as it is difficult to control the factors like smoking, drinking and emotional undulation, only the overall condition can be generally reflected. In their research, Akselrod discovered in 1981, that the power spectra of HRV features derived from ECG signals obtained by non-invasive measuring which have further been processed by Fourier transform can reflect the control of sympathetic and parasympathetic nervous systems on physiological mechanisms. By doing spectrum analysis on R-R interval, it is possible to observe three noticeable spectral peaks, which are high frequency (HF), low frequency (LF) and very-low frequency (VLF) and which are reflected in HRV. In 1996, the American Heart Association (AHA) set new criteria for the bandwidths of HRV, which are 0.15-0.4 Hz as HF bands, 0.04-0.15Hz as LF bands and 0.003-0.04Hz as VLF bands. The LF represents sympathetic nervous activity, HF for parasympathetic nervous activity and VLF for temperature regulation, Rennin-Angiotensin-Aldosterone System (RAA System) and part of parasympathetic nervous functions. HRV is controlled by autonomic nervous system (ANS), whereby to maintain the body balance, and the variance in heartbeat interval is affected by sympathetic and parasympathetic nervous systems; hence, the study of HRV can serve as an index for the autonomic nervous system [2].

HRV analysis, as a physiological index, began to catch doctors' attention in recent years. It has been used to monitor the regulation of human body by the autonomic nervous system, also served as a reference index in the death rate due to cardiovascular diseases. The commonly seen indices for death due to cardiovascular diseases are the LF rise of sympathetic and partly parasympathetic nervous regulations, and the HF decrease of parasympathetic nervous regulation. When the diagnosis is a likely coronary artery disease, it often further takes cardiac catheterization. Patients are often seen with emotions like anxiety and unease and become anxious or fearful for the dangers and side effects that may occur during the examination or the uncertainty of examination results, where they need appropriate caring guide to mitigate the anxiety. These emotions are potential pressures to patients of cardiovascular diseases. In anxiety, the rate of occurrence of sudden heart diseases also increases while the HRV is seen decreasing. Research showed that in patients of Myocardial Infarction, the emotions of anxiety and panic is associated with HRV decrease, and that both anxiety and melancholy decrease HRV and are related to the increase in risk of cardiovascular death in patients of coronary artery disease and irregular heartbeats. When Zen meditation is used to intervene

in the measurement of characteristic anxiety scale and the changes of HRV analysis, the findings were that in sufferers of high-degree anxiety who underwent Zen meditation training and concentrated, the anxiety decreased the scores on the scale, the scores on characteristic anxiety scale were reversely related to slow alpha wave ($r=-0.482, p<0.05$) and positively related to HF index of HRV ($r=0.555, p<0.01$), but were not significantly related to LF and LF/HF ratio ($r=0.305, p<0.17$). It was reported that before cardiac catheterization, both contextual anxiety score and characteristic anxiety score in men were lower than in women, where the characteristic anxiety scale and HRV were significantly different in terms of gender ($t=-2.44, p<0.019$), while contextual anxiety scores were insignificant statistically. Also, in the gender differences in HRV analysis, it was seen that SDNN of HRV by time domain methods and HF and LF indices in women were all significantly greater than in men ($p<0.05$). Mental stress causes unstable emotions like negative one such as dysphoria, anger and discontentment, which affect the autonomic nervous system, stimulate the changes in sympathetic and parasympathetic nervous activities and are reflected in heart rate. In the research in short-period R-R interval and spectra analysis of cardiovascular systolic pressure, which explored the effects of psychological stress on autonomic nervous system and circulation system, the finding was that when concentrating, the subjects had higher-than-usual energy in the LF bands. That same study observed the correlations between skin temperature, vascular changes and sympathetic nervous activities using the indices of HRV, changes in skin temperature and vasomotor nervous activities of peripheral vessels, in the condition of mental stress and biofeedback. It was believed in that document that the degree of sympathetic nervous activities could serve as an index for durability to pressures. It was reported that under mental stress, HF energy dropped greatly and LF/HF noticeably rose and was significant statistically ($P<0.05$); when mental stress decreased, the LF/HF ratios lowered notably and was significant statistically ($P<0.05$). It is thus clear that when mental stress is strong, sympathetic nervous activities would increase by far, which can serve as an index for the assessment of the condition of mental stress, further to assist in clinic testing and diagnosis.

Now we talk about the recognition of facial emotions. Ekman and Friesen were the first to develop Facial Action Coding System (FACS), which defined 44 different Action Units (AU) by the principle that man pulls different facial muscles when making different expressions [3]. As Fig.1 shows, Fig.(a) is action units of upper face and some combinations of action units, and Fig.(b) action units of lower face. By combining certain action units, human facial expressions can be depicted. The FACS theory has been applied in most research to identify facial expressions. While applying it, a study made a difference in the capturing of facial features and the selection of recognition and sorting methods [4]. It divided the facial features in constant features and transient ones; the former were the changes in eyes and mouth and the latter were the wrinkle changes in between the

brows, cheeks and upper mouth edges. That study manually set the initial positions for these feature points, followed by tracking the changes of these points and then recognizing them with Artificial Neural Network [4]. Another study used Optical flow to track the movement of multiple feature points, which were set in a first image of a sequence of images. The characteristic vectors that were produced combining the feature point displacements in the first image and the last one were used to identify different action units and different facial expressions [5].

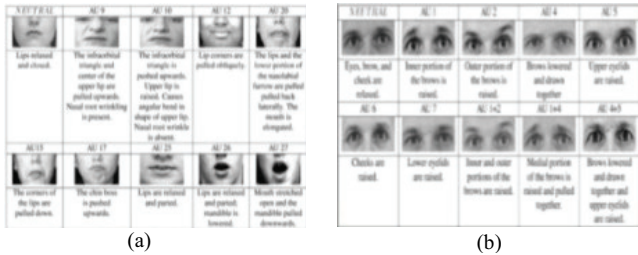


Figure 1. Facial Action Units [5]

Feature capturing involves two aspects: the holistic approach and the local approach. The holistic approach is about first obtaining feature areas or feature points, whereby to detect deformation or movement, such as optical flow and characteristic tracking [6]. The local approach divides human face in several key blocks and then extracts characteristic information from them for determining the action [7].

This paper includes four chapters, first chapter is introduction, second is documentary review and exploration, the third is methodology and steps, we introduce hardware architecture and our methods, the fourth is experiment results, and the last is conclusions.

III. METHODOLOGY AND STEPS



Figure 2. Diagram of system framework

The development of emotion recognition system plays an important role in the research of psychological disorders and health care. This system will be discussed in separate sections: software and hardware. The hardware comprises an embedded system as the core, which works with Webcam to capture facial images and with Blue-tooth electrocardiograph (BtECG), to measure physiological signals. The software incorporates Webcam driver, Fuzzy Inference method, HRV analysis, facial detection and the techniques of eye and mouth state recognition.

In the system, the Webcam captures facial images; for reference values used to determine by the Fuzzy Inference method, thirteen features, as follows, are used: distance between eyeballs, number of changes in eyeball height, the extent to which the mouse opens, SDNN, Standard deviation

of the averages of NN intervals in all 5-minute segments of the entire recording (SDANN), RMSSD, Number of pairs of adjacent NN intervals differing by more than 50 ms in the entire recording (NN50 count), NN50 count divided by the total number of all NN intervals (pNN50), Mean RR intervals (RRI-Mean), Standard deviation of differences between adjacent RR intervals (SDSD), Standard deviation of all normal to normal intervals index (SDNN index), VLF, LF, HF and analysis of LF/HF power. The job of embedded system is processing the measured HRV features and the imaged the Webcam sent, and detecting user's facial positions and identifying the state of the eyes and the mouth by facial detection and eye-and-mouth state identification. In this paper to combine the Fuzzy Inference method to distinguish the differences in degree of emotion based on the features of the eyes, the mouth and HRV. The system outputs the results of determination to a screen or informs relevant persons by texting. Elaboration will now be made on the software.

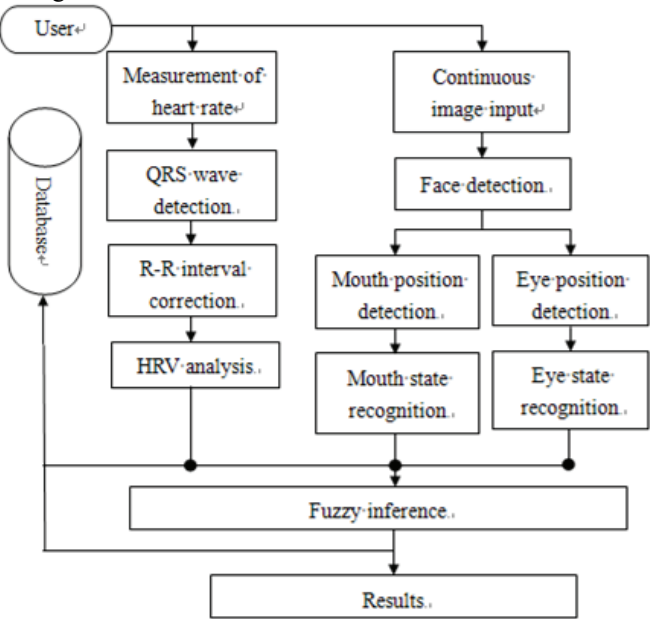


Figure 3. Flow chart for emotion recognition system

The flow chart of this paper is shown in Fig.3. To begin with, relevant data are downloaded to the embedded system via a USB device for use. That is followed by developing USB Webcam driver codes in the embedded system to enable the WINCE platform to drive Webcam to capture images, which will be saved in the embedded system. Again, this is followed by pre-process of the user's image, which aims to eliminate the influences of indoor light and noises in order to enhance the accuracy of capturing. A technique of graph identification is used to detect the state of the eyes and the mouth and then a BtECG is used to transmit physiological signals to the embedded system. The QRS-wave detection is employed to make correction to R-R interval before analyzing HRV by time domain methods and frequency domain methods with the features. When characteristic identification has been conducted, the fuzzy inference method is applied to combine the HRV features and the features on images, with their

degrees distinguished. Lastly, based on different degrees, the system analyzes and determines the user's emotional state. In respect of above flow, the research method is divided in below steps:

A. Facial detection

- (1) Preprocessing: the system employs an image pro-process, of which the object is to eliminate noise interference in the image and to adjust brightness as well as remove the noise.
- (2) Skin region detection:

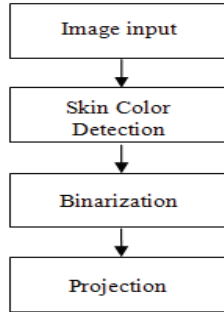


Figure 4. Flow chart for facial detection

Detection of facial and eye position is the first and important step of eye status recognition. There are many detection methods of facial position; but this paper employs facial complexion in detection. Also in this paper, the images having been processed by binarization are taken to locate the face after horizontal and vertical projection. The method of complexion detection uses the information on complexion to capture the face in an image, discerning it from the background. An image is transformed from the RGB space to the $YCbCr$ space.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1687 & -0.3312 & 0.5 \\ 0.5 & -0.4183 & 0.0816 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$60 < Y < 255 \quad (2)$$

$$25 < C_b < 0 \quad (3)$$

$$10 < C_r < 45 \quad (4)$$

As the complexion has a fixed zonal distribution in the $YCbCr$ space, this paper sets the pixels that distribute in such zone as 1 and the values in this zone as 0. As the face occupies the largest one of a great number of blocks taken in the image by complexion, the method of projection is used to identify the complexion block of biggest area. The area is gradually closed by morphological operations algorithm, and the separate pixels are conjoined through expansion; further, the areas that are weakly connected are separated by erosion to obtain binary images.

B. The Detection of Feature Point Position

(1) Eye detection:

The length of human face is about 1.2 times of the width.

According to the facial features proportion, when starting from the top of the face, the pupil will be at the 4/12 to 7/12 of the face length. Starting from the left of the face, the pupil will be at the 2/16 to 15/16 of the face width. We find the pupil in this area. Compared with the nearby area, the pupil is of darker color. Thus we use the dynamic method to find the critical value. First calculate the grayscale histogram of this area. Since the pixels of the pupil is within the 1% of the darkest area, we calculate from the grayscale of 0, accumulate to the 1% of the total pixels of this area and set the value of this pixel as the critical value. When the pixel value of the area is smaller than the critical value, set it as 1, otherwise set as 0, as shown in Fig. 5.

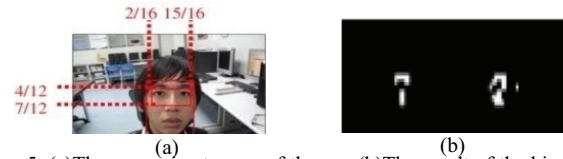


Figure 5. (a) The approximate area of the eye (b) The result of the binarization of the dynamic critical value of the pupil area

Then we separate the eye area into right and left two sections. Raise the left pupil as an example, in the process of binarization of the pupil, it may also binarize the eyebrow and hair near the eye. This will affect our action of finding the position of the pupil. However, because the eyebrow and hair are above the pupil, our algorithm may search the first pixel of 1' in a bottom up way. And then use this pixel as starting point, which, together with two pixels on the left and right, and 5 pixels above will generate a 5*5 area. And then calculate the center of gravity of all the pixel value 1's pixels in this area. We find that this center of gravity is the position of the pupil. For the same reason, we can find the right pupil with the same method. And then use the distance(d) of the two pupils as in Fig. 6, follow the relationship of the human face features' proportion to generate the area of the eye. The eye area is at the place where the pupil is the center, go up and downwards 1/6 distance, left and rightwards 1/3 pupil distance area.

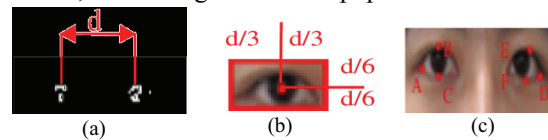


Figure 6. (a) Pupil distance d (b) the eye area detection result (c) the feature point of the outer corner of the eye.

After finding the eye area, we utilize the pixel center of gravity obtained by the above 5*5 area. As this pixel center of gravity is always inside the pupil area and the grayscale value is similar to that of the other pixels in the pupil, we use the Region Growing method to use the pixel center of gravity as seed to grow complete pupil area. In this way, we will know the coordinates of the uppermost pixel point and that of the bottommost pixel point.

When we find the top and bottom threshold of the pupil, we will find the outer corner of the eye to be the other feature point for identification. Use the top and bottom feature points as the upper and lower bounds along the y axis. Then use the pixel center of gravity as the starting point and seed along the

x axis. Similarly we use the Region Growing method along the outer corner of the eye orderly. When scanning the connecting zone of the pupil and the white of the eye, due to the fact that the white of the eye is not the pupil area, the grayscale may change greatly. Record this pixel point as the first point, then use the next pixel point, i.e. the pixel point inside the white of the eye as seed for processing. Similarly when scanning the connecting zone of the white of the eye and the skin; the grayscale may change greatly. Record this pixel point as the second point. Repeat these processing steps till the boundary of the eye area. Finally find the outermost point of the points we record. This is the feature point of the outer corner of the eye we need, as shown in Fig. 6(c).

(2) Mouth detection:

When acquiring the feature point of the mouth, we first use the facial features proportion to generate the approximate area of the mouth. The position of the mouth will be at the 7/10 to 9/10 face length and 2/6 to 5/6 face width area when starting from the top of the face, as shown in Fig. 7. As the color of the lip is redder than other skin, we use this information to obtain the feature point of the mouth. First convert the human face image from RGB color space to HSV color space. In this way we can separate the color and the brightness to avoid the influence from the outside light source. Then use (1) to transform the H (hue) value in the HSV color space. The equation is as follows:

$$f(h) = \begin{cases} 1 - \frac{(h-h_0)^2}{w^2}, & |h-h_0| \leq w \\ 0 & |h-h_0| \geq w \end{cases} \quad (5)$$

Where the h is the h value of the mouth area, h_0 is the offset value defined according to the hue. Here we set it as 1/3. The w is set as 0.60. This equation can be viewed as use the h_0 as center and use w to filter the area that can be accepted. Farther the distance between h value and h_0 of the mouth area, smaller the value will be. The value of f (h) is between 0 and 1. Then we amplify the f (h) value of the mouth area. Round off the f (h)*100 value into integer and amplify it 100 times to produce histogram as shown in Fig. 8. We found from the observation that when the value of f (h)*100 is between 50 and 100, there is a peak as the skin. There is a point on the left-hand side with a large slope. When the gradient is gentle, it is suitable to be the critical value to separate the mouth and the skin. Thus if the f (h)*100 value is smaller than this critical value, it is the mouth area. Set the pixel value as 1, otherwise set it as 0. Binarize the image as shown in Fig. 9.



Figure 7. The approximate mouth area

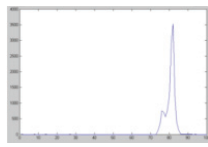


Figure 8. f (h)*100 histogram



Figure 9. Mouth area binarization result.

As the mouth has the horizontal edge direction tendency, we use the horizontal Sobel wave filter to generate the horizontal direction edge of the mouth area and do the binarization. Then search from left to right to find the point

where the first mouth binarization and Sobel binarization pixel value both are 1. This is the feature point G. Search from right to left to find the point where the first mouth binarization and Sobel binarization pixel value both are 1. This is the feature point I. From the horizontal direction center point of the left and right feature points, 2 pixels on the right and left, there will be 5 horizontal pixels in total. Search top-down to find the point where the first in these five horizontal pixels, the mouth binarization and Sobel binarization pixel value both are 1. This is the feature point H. Search bottom- up to find the point where the first in these five horizontal pixels, the mouth binarization and Sobel binarization pixel value both are 1. This is the feature point J as shown in Fig. 11.



Figure 10. Mouth Sobel binarization result

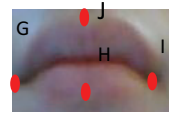


Figure 11. Mouth feature points

C. Heart Rate Variation Analysis

Regarding the studies of the periodicity of the heart beat, they mostly use the continuous R-R interval as the continuous period of the heart beat. The R wave of the electrocardiogram is the most commonly used in the heart rate variation analysis. By calculating the time interval between the RR, we get a series of numbers. The heart rate variation analysis usually excludes the Arrhythmia case. It focuses on the slight heart rate variation of the normal rhythm for study.

First, we stick the sticky thread on the user's left and right wrist, and then use the BtEEG to measure the physical signal of the user. The measuring time is 5 minutes. Transfer the physical signal to the mosaic system through the BtEEG. To reduce the interference of the heart rate variation result, we keep the environment quiet. In this paper we use the method proposed by Willis J. Tompkins to detect QRS wave. This method follows the frequency range of the signal to filter out the noise to get the QRS wave. We can find the position of the R wave through the QRS wave. Fig. 12 is the flow chart of using the Willis J. Tompkins' method detecting QRS wave.

- Band-pass integer filter: As the center frequency of QRS wave is about 17Hz, we use a low-pass filter (25Hz) and a high-pass filter (15Hz) to form a band-pass integer filter and keep the frequency between 15 and 25Hz. In this way, we can get a clearer QRS complex wave and reduce the T wave substantially.
- Differential: After the signal passes the band-pass integer filter, we can use differential to make features of QRS wave more high-lighted. After differential, we can easily distinguish the R wave.
- Square: In order to convert the signal into the positive and make high-frequency part of the QRS wave more obvious, we use square to make it stand out.
- The acquisition of R wave's position: We get the peak value of R wave by adjusting the threshold dynamically. However, someone's T wave is higher than his R wave or only slightly

lower than his R wave. We have to filter out T wave before catching R wave, and then search the peak value of R wave. In this way, we can find the next R wave and obtain the R-R interval.

However, R-R interval will change because of age or certain diseases of the user. Therefore, this program will use the above method to find the R-R interval, and modify it by using the methods proposed by [7]. The specific steps are as follows:

1. Expand the above R-R interval*0.6 to the left of R point and R-R interval * 0.3 to the right of R point. Repeat this step and select 20 heartbeats randomly.
2. Use the method of Signal averaging to make 20 heartbeats' electrocardiogram point at R point. After getting the cumulative average, we should calculate the related feature value for identification.

At present, there are two kinds of methods on analysis of heart rate variability, namely time domain analysis and frequency domain analysis. Time domain analysis is to record the statistical analysis result of R-R interval between two neighboring heart rates, while frequency domain analysis is to use the fast Fourier transform to convert the R-R interval from time domain into frequency domain for Power Spectral Density (PSD) analysis. This program will analyze time domain and frequency domain, and make time domain analysis through the following features

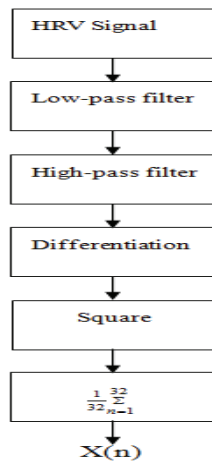


Figure 12. The flow chart of using Willis J. Tompkins' method to detect QRS wave[8].

1. SDNN: Standard deviation of Sinus rhythm, with the clinical significance to estimate the overall heart rate variability.
2. SDANN: Standard deviation of the interval average of all normal heartbeats every five minutes.
3. RMSSD: The square and the average square root of neighboring R-R interval difference.
4. NN50 count: The count of the time interval of each pair of neighboring normal heart rates that is more than 50 milliseconds.
5. pNN50: NN50 count divides all the normal heartbeat intervals in the ECG.
6. RRI-Mean: The average RR interval every 5 minutes.
7. SDSD: Standard deviation of differences between adjacent

RR intervals.

8. SDNN index: Calculation of the standard deviation of normal heartbeat every 5 minutes and then get the average value.

This program adopts the above eight features to analyze time domain, and uses the Bluetooth ECG machine of Yang Yin Inc. to conduct a five-minute measurement of the user.

Frequency domain analysis:

We use the fast Fourier transform to convert the R-R interval from time domain into frequency domain and then display it in the way of Power Spectral Density (PSD). According to the size of the different frequency, Power Spectral is divided into total power, ultra-low frequency, low frequency, high frequency and low frequency vs. high frequency. This program will focus on analysis of ultra- low frequency, low frequency, high frequency and low frequency vs. high frequency.

D. Feature Identification

After getting the position of various features and the features of heart rate variability, we will make the feature acquisition calculation. This program touches upon the values of thirteen features, namely, the distance between eyeballs, the size of open mouth, times of eyeball height changes, SDNN, SDANN, RMSSD, NN50 count, pNN50, RRI-Mean, SDSD, SDNN index, ultra-low frequency, low frequency and high frequency vs. low frequency. As the above feature values will vary in different emotional states, the distance between eyeballs and times of eyeball height changes will vary with the emotion. We can analyze the distance between eyeballs through (2):

$$eyed = \frac{|B - C|}{d} \quad (6)$$

As is shown in Fig. 7 (c), $|B - C|$ is the vertical distance from B to C of the feature point in the right eye, while "d" is the distance between two pupils in Fig. 7(a). In different emotional states, two eyes do the same action, so we make judgment based on the right eye. We analyze the mouth openness through the Euler distance, which is the vertical distance from H to J in Fig. 11. Accordingly, there are also different conditions on heart rate variability, for example, when depressed, SDNN value will be lower than usual, and the high frequency energy will plummet. Low-frequency will increase remarkably when there is huge mental stress, and it will decrease significantly when the mental stress is decreased. This program will analyze four emotional states, normal, happy, sad and angry through Fuzzy Inference method. We will use the thirteen features to find out the attribution between the input and output variables. First, we will set up the semantic formula of fuzzy inference on input and output variables:

- (1) The distance between eyeballs: If it is smaller than 0.125 cm , it is diagnosed as mild eye openness(S), from 0.1 to 0.15 cm as moderate eye openness (M), and higher than 0.125 cm as high eye openness (L), whose membership function is μ_{eyed}
- (2)Times of the height change of eyeballs: If it is less than 10 times, it is small amount of change (S), 5 to 20 times is

moderate changes (M), and more than 10 times is large amount of change (L), whose membership function is μ_{eyeH} (3)The frequency: if it is higher than 0.15, it is called high frequency, from 0.05 to 0.15 is called low frequency, and lower than 0.05 is called ultra-low frequency, whose membership function is $\mu_{frequency}$ (4)SDNN: the attribution function is μ_{SDNN} (5)SDANN: the attribution function is μ_{SDANN} (6)RMSSD: the attribution function is μ_{RMSSD} (7)NN50 count: the attribution function is $\mu_{NN50count}$ (8)pNN50:the attribution function is μ_{pnn50} (9)RRI-Mean: the attribution function is $\mu_{RRI-Mean}$ (10)SDSD: the attribution function is μ_{SDSD} (11)SDNN index: the attribution function is $\mu_{SDNNINDEX}$ (12)Low frequency vs. high frequency: the attribution function is $\mu_{LF/HF}$ (13)The mouth openness: If the distance from H to J is less than 1.48 cm, it is slight openness (S), from 1 to 2.21 cm, it is moderate openness (M), and larger than 1.48 cm is large openness (L), whose attribution function is μ_{mouth} .

In the following construction of fuzzy rules, we can use semantic fuzzy rules obtained by the doctor or through learning, and finally get the fuzzy sets. Then we use Center of Gravity Method for defuzzification. We can also distinguish the four emotions into three levels, that is slight(S), moderate(M) and large(L). Fig. 13 is the attribution function of the above features and each kind of emotion.

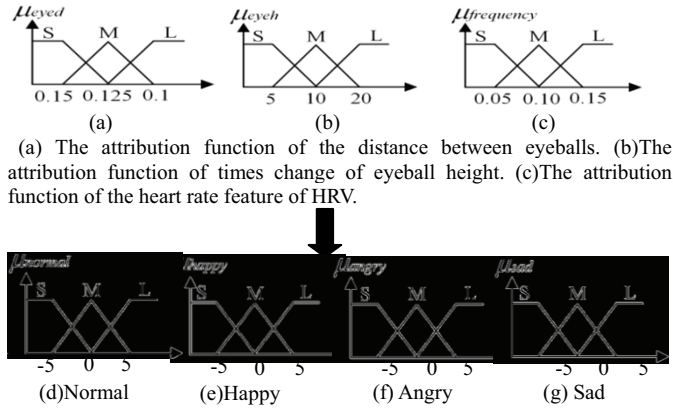


Figure 13. The flow chart of emotional state inference.

E. The Output Result

Based on the input image and the features of heart rate variability, this program shows the emotional states of the current users by using inference results of Fuzzy Inference, that is, normal, happy, sad and angry. What's more, this program stores the image, features of heart rate variability and emotional states into the database, which can provide the reference for doctors in case of medical treatment in the future.

IV. EXPERIMENT RESULTS

The results of this paper, we propose an intelligent method to construct the emotion recognition system for testing and discussion its recognition rate. Herein, the CCD

camera is QuickCam™ Pro 4000 of Logitech. The image resolution is 320×240, and the camera capture 30 images per second. Because the measured time is five minutes, we can capture 9000 images. Meanwhile, the paper uses embedded system for testing, and its operating system is Windows Embedded CE 6.0 R2, the RAM size is 1024MB. The distance from human face to CCD camera is 40cm~60cm, the speed of recognition spends about 0.4~0.6sec. In addition, the part of measuring heart rate, we use Bluetooth ECG device to do measurement, which is produced by yang yin co. The measuring environment is in the laboratory, and keeps quiet for five minutes measurement; we conducted heart rate and face test for 20persons.

Fig. 16 and 17 show the results of face detection; however, because of light and noises we still have mistakes in human face detection. Fig. 16 shows successful cases in human face detection and Fig. 17 shows failure detections. And the failure is resulted from skin's color range and lights.



Figure 14. Successful cases in human face detection



Figure 15. Failure case in human face detection

After detecting human face, we proceed to extract facial features. Fig. 18 shows the result of the extraction of feature points.



Figure 16 shows the result of the extraction of feature points.

Next, to extract the feature of heart rate, we extract 20 people's respectively extract 4 emotions electrocardiogram for each 20 persons, Fig. 19 show the electrocardiogram of (a)normal, (b)happy, (c) angry, (d) sad, respectively.

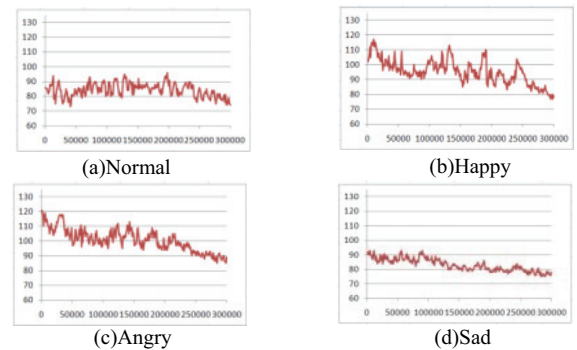


Figure 17. The electrocardiogram of four emotions

After obtaining above features, the paper use fuzzy inference method to judge emotion state. In fuzzy inference, we use 13 features, and divide them into three levels. They are slight(S), moderate (M) and large (L), respectively. The system is multi-input and single-output, and the emotion is divided into sad, angry, happy, normal, after deduction. We

also divided 4 emotions into three levels. Because there are 13 features in the system and infer a emotion state, total rules have 3^{14} rules, after reducing, the total rules remain 3^{10} rules. For instance:

R¹:if eyed is S, eyeh is S, frequency is M, SDNN is M, SDANN is S, RMSSD is S, NN50 count is S, pNN50 is S, RRI-Mean is S, SDSD is M, SDNN index is M, $\frac{LF}{HF}$ is S, mouth is S Then normal is S.

R²:if eyed is M, eyeh is M, frequency is M, SDNN is M, SDANN is S, RMSSD is M, NN50 count is M, pNN50 is S, RRI-Mean is S, SDSD is M, SDNN index is M, $\frac{LF}{HF}$ is S, mouth is M Then happy is M.

R³:if eyed is L, eyeh is L, frequency is L, SDNN is M, SDANN is L, RMSSD is M, NN50 count is L, pNN50 is L, RRI-Mean is S, SDSD is M, SDNN index is M, $\frac{LF}{HF}$ is M, mouth is M Then angry is M.

Table I shows the average successful rate of mouth and eyes detection is 82.03%.The success rate of our system is 84.31%.

Table I: the success rate of facial feature points detection

	The number of detection of eyes error	The number of detection of mouth error	Success rate
video 1	818	886	81.06%
video 2	952	1159	76.54%
video 3	1102	897	77.79%
video 4	1452	1220	70.31%
video 5	891	925	79.82%
video 6	751	862	82.07%
video 7	991	1123	76.51%
video 8	729	982	80.98%
video 9	902	1163	77.05%
video10	685	746	84.10%
video11	859	957	79.82%
video12	745	712	83.81%
video13	812	892	81.07%
video14	688	657	85.06%
video15	901	972	79.19%
video16	742	803	82.83%
video17	992	902	78.95%
video18	817	933	80.56%
video19	700	721	84.21%
video20	1102	801	79%
Average successful rate 82.03%			

V. CONCLUSIONS

We provide an emotion recognition system by using eye's close/open condition, mouth close/open condition, physiological signals, and the method of Fuzzy Inference and apply emotion recognition system in the embedded system. This system is light, so has less spatial constraint. However, there are still parts required improvements in this emotion recognition system. For example, we hope natural light can be used to replace external light, so they obtained eyeball images will not so easily be influenced by reflection points and image capturing will be more accurate. Another solution is to find a webcam that has higher resolution and can filter some external image noises. However, problems regarding how to reduce dynamic distribution but with clear captures on the eyes or how to emphasize pupil's features should be dealt with by designing other appropriate image processing rules. In addition, we want to improve QRS detection in heart rate detection. In facial detection, the average successful rate is 82.03%, and the success rate of our system is 84.31%.

In the future, we expect that in the parts of human face detection, features capture and heart rate detection, we can process with better algorithm (for example, the color space of light's images and capturing more featured areas), establish a database to collect more reference materials.

REFERENCES

- [1] G. Zhou, Y. Zhan and J. Zhang, "Facial expression recognition based on selective feature extraction," *International Conf. on Intelligent Systems Design and Applications*, Vol. 2, pp. 412-417, 2006.
- [2] T. Salonidis, P. Bhagwat, L.Tassioulas, R.LaMaire, "Distributed topology construction of Bluetooth personal area networks," *INFOCOM. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp.1577-1586, April 2001.
- [3] P. Ekman and W.V. Friesen, "The facial action coding system: A Technique for The measurement of facial movement," *San Francisco, Consulting Psychologists Press*,1978.
- [4] J. F. Cohn, A. J. Zlochower, J. J. Lien and T. Kanade, "Feature-Point tracking by optical flow discriminates subtle Differences in facial expression," *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara,Japan, pp. 396-401, 1998.
- [5] Ying-li Tian, T. Kanade and J.F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.
- [6] Jiapu Pan, and W. J. Tompkins, "A Real-time QRS Detection algorithm," *Trans. On Biomed. IEEE*, Vol. 32, No.3, pp.230-236, 1985.
- [7] Fleet, R., Lavoie, K., & Beitman, B. D., "Is panic disorder associated with coronary artery disease? A critical review of the literature," *Journal of Psychosomatic Research*, Vol.48, pp.347-356, 2000.
- [8] Murata, T., Takahashi, T., Hamada, T., Omori, M., Kosaka, H., Yoshida, H., "Individual trait anxiety levels characterizing the properties of zen meditation," *Neuropsychobiology*, Vol.50, pp.189-194, 2005.