A Non-invasive Facial Visual-Infrared Stereo Vision Based Measurement as an Alternative for Physiological Measurement

Mohd Norzali Haji Mohd[1,2], Masayuki Kashima[1], Kiminori Sato[1], Mutsumi Watanabe[1]

[1] Department of Information Science and Biomedical Engineering, Graduate School of Sciences and Engineering, Kagoshima University, Korimoto 1-21-40, Kagoshima, 890-0065, Japan.
[2] Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, University Tun Hussein Onn Malaysia (UTHM),86400, Parit Raja, Batu Pahat, Johor, Malaysia.

Abstract. Our main aim is to propose a vision-based measurement as an alternative to physiological measurement for recognizing mental stress. The development of this emotion recognition system involved three stages: experimental setup for vision and physiological sensing, facial feature extraction in visual-thermal domain, mental stress stimulus experiment and data analysis and classification based on Support Vector Machine. In this research, 3 vision-based measurement and 2 physiological measurement were implemented in the system. Vision based measurement in facial vision domain consists of eyes blinking and in facial thermal domain consists 3 ROI's temperature value and blood vessel volume at Supraorbital area. Two physiological measurement were done to measure the ground truth value which is heart rate and salivary amylase level. We also propose a new calibration chessboard attach with fever plaster to locate calibration point in stereo view. A new method of integration of two different sensors for detecting facial feature in both thermal and visual is also presented by applying nostril mask, which allows one to find facial feature namely nose area in thermal and visual domain. Extraction of thermal-visual feature images was done by using SIFT feature detector and extractor to verify the method of using nostril mask. Based on the experiment conducted, 88.6% of correct matching was detected. In the eyes blinking experiment, almost 98% match was detected successfully for without glasses and 89% with glasses. Graph cut algorithm was applied to remove unwanted ROI. The recognition rate of 3 ROI's was about 90%-96%. We also presented new method of automatic detection of blood vessel volume at Supraorbital monitored by LWIR camera. The recognition rate of correctly detected pixel was about 93%. An experiment to measure mental stress by using the proposed system based on Support Vector Machine classification had been proposed and conducted and showed promising results.

1 Introduction

Various methods for internal state measurement such as mental stress have been carried out which utilize the changes of physiological through electroencephalography (EEG), blood volume pulse (BVP), heart rate variability (HRV), galvanic skin response (GSR) and electromyography (EMG) measurement [2]. However it requires the individuals to wear or touch electrodes or sensors. On the other hand, physical signals for measuring stress include eye gaze, pupil diameter, voice characteristic and face movement and these quantities are measured invasively by the use of expensive instruments.

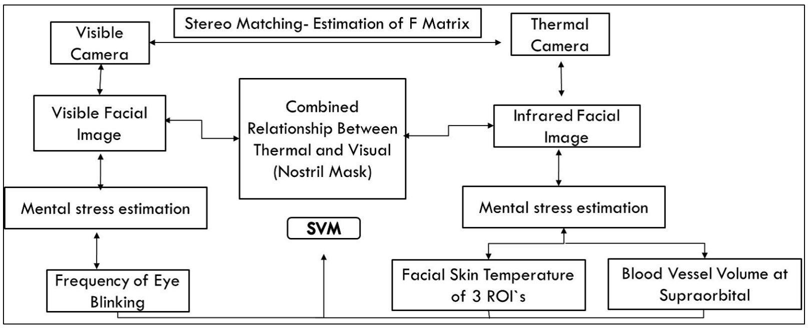

Fig. 1. Overall System.

1.1 Problems and related work

Face recognition system based on visual images has reached significant level of maturity with some practical successes. However, the performance of visual face recognition may degrade under poor illumination conditions, especially for subjects of various skin color and the changes in facial expression. The use of infrared in face recognition allows the limitations of visible face recognition to be solved. However, infrared suffers from other limitations like opacity to glasses. Hence, multi modal fusion comes with the promise of combining the best of each modalities and overcoming their limitations [4]. Dvijesh Shantri et. al [10] proposed a novel framework for quantifying physiological stress at a distance via thermal imaging. The method captures stress-induced responses on the perinasal area that manifest as transient perspiration. It is based on morphology and spatial isotropic wavelets. The limitation of this propose perinasal imaging is it depends on an expensive sensor, cannot operate if the perinasal area is covered and requires support from reliable face tracker.

1.2 Key Contribution area

In this paper we introduce an integrated non-invasive measurement via imaging techniques. Our aim is to propose vision-based measurement as an alternative to physiological measurement. The fusion of physiological vision-measurement from thermal IR and visual is shown in Figure 1. Our research consists of three vision based physiological measurements which are eye blinking from visual sensor, skin temperature of 3 ROI‘s and blood vessel volume at supraorbital from thermal IR camera. The primary physical measurement in detecting mental stress is heart rate variability [2] and Salivary amylase level [3] as proof by earlier researcher. The normal heart rate ranges from 60-100 bpm. In other hand, Salivary amylase level increased significantly and is suggested as the better index of mental stress [3]. Salivary amylase with level more than 60 $KU \backslash L$ is considered to have mental stress. Both of the measurement was used as a ground truth measurement in our research. We will also publish the data set consisting the proposed vision measurement.

In this paper, we also propose a new calibration chessboard attach with fever plaster to locate calibration point in stereo Visual-Thermal view. A new method of integration of two different sensors for detecting facial feature in both thermal and visual is also presented by applying nostril mask, which allows one to find facial feature namely nose area in thermal and visual domain.

Fig. 2. Screening Station.

2 Methodology

2.1 Registration of IR and Visible

We propose three ways to compute the relationship between Visible-IR camera as below.

Registration of IR-Visible by using fever plaster calibration board. The relative position between the IR and visible cameras is calibrated by using the special calibration board. With some small adjustment and preparation where cold fever plasters are attached to the back of the calibration points on a chessboard, the calibration points can be reliably located in thermal IR and visible domain[16]. One of the common strategies to simplify correspondence problem between IR and visible domain is to exploit epipolar geometry (Fig.4). The epipolar line is the intersection of an epipolar plane with the image plane. All epipolar lines intersect at the epipole. An epipolar plane intersects left and right image plane in epipolar lines, and defines the correspondence between lines [6]. The relative position between thermal-visible stereo cameras is calculated using calibration board. The origin of the world coordinate system is defined to be on the calibration board.

Fig. 3. Propose Calibration chessboard.

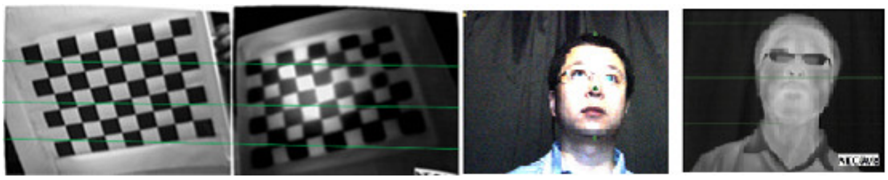

Fig. 4. Detected F matrix and epipolar lines from Th-Vi calibrated images.

The comparison between traditional heated chessboard and the proposed calibration board is cheap in comparing to the other calibration board made from polished copper plate coated with high emissivity paint or calibration rig. Even though heating is required in both method, it is difficult to get an even coverage which the pattern that can last longer than 10 minutes. We also propose different ways to compute the relationship between Visible-IR camera.

SIFT feature matching. Facial feature in both thermal and visible that can be detected in view is nostril part (Fig. 5). SIFT features are extracted from images to help in reliable matching between different views in same object [9]. The extracted features are variant to scale and orientation, and are highly distinctive of the image. They are extracted in four steps. The first steps computes the location of potential interest points in the images by detecting the maxima and minima of a set of Difference of Gaussian(DoG) filters applied at different scales all over the images. Then, these location are refined by discarding points of low contrast. An orientation is then assigned to each key point based on local image features. Finally, a local feature descriptor is computed at each key point. This descriptor is based on the local image gradient, transformed according to the orientation of the key point to provide orientation invariance. Every feature is a vector of dimension 128 distinctively identifying the neighborhood around the key points.

At first, the nostril area in thermal IR is detected using our pair calibrated Thermal-Visible camera[15]. Then, the face part in visible domain is detected and SIFT feature point is calculated in both domain. The feature matching is done and the correct matching is recorded in Table 1.

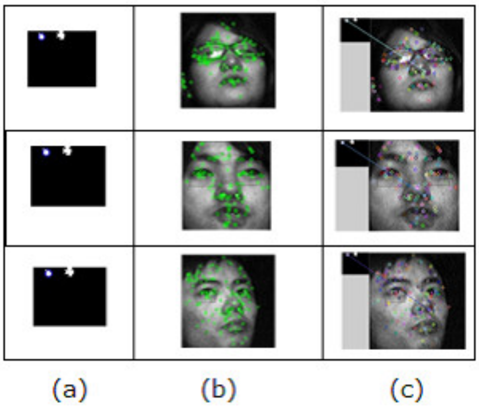

Fig. 5. (a)Feature point (nostril mask in thermal), (b)Feature point(facial in visual) (c) Sift feature matching.

Stereo matching: estimation of F matrix The correspondence problem remains of central interest in the field of image analysis. In the case of uncalibrated cameras, establishing correspondence between cameras is crucial. In our research, the matching of feature points correspondence between stereo pairs of visible and infrared facial is done as one of the matching strategy. Our focus is only on frontal faces as the stimulus is located in front of correspondent. One of the common strategies to simplify correspondence problem between IR and visible domain is to exploit epipolar geometry. Our system is based on uncalibrated images taken from thermal and visual camera (Fig. 6) .The reason of this step is to show the correspondence and location of facial images taken from thermal IR and visible camera. By knowing the point correspondence in two images we can compute and estimate F matrix. SURF feature point is used to extract and match feature points correspondence. The outliers are then removed by using epipolar constrain. Then the images are being rectified on F matrix.

$$m'^T F m = 0 \tag{1}$$

Lastly, transform points are visualized together using composites image.

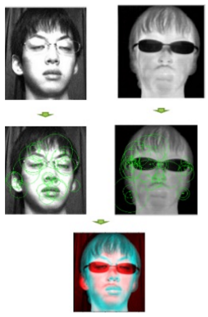

Fig. 6. Estimating F matrix, remove outliers using epipolar constraint and visualize rectified images by composites image.

2.2 Facial Feature Extraction in Visual and Thermal Domain

Frequency of eye blinking from visible domain. At first we detect the face using Viola and Jones' boosting algorithm and a set of cascade structure with Haar-like features[17]. The first 30 frames of detected face that shows the location is marked as template. The edge image of the face shows the location of the eye more precisely because of the contour of the eyes area. Eyes regions are detected by calculation of integral image rectangular filter.

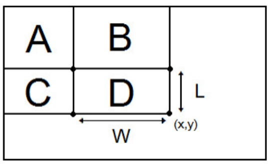

Fig. 7. Calculation of the integral image rectangular filter.

$$S_r = (ii(x,y) + ii(x-W, y-L) - ii(x-W, y) - ii(x, y-L)) \tag{2}$$

$S_r$sum of the pixel values of the rectangle D is obtained by calculating the difference between the sum of the 4-points. The filter is minimum. When the location of eyes region is detected, the location of the iris is determined where there are the most black pixel in the horizontal axis. Fig.8 below shows the steps in locating iris area.

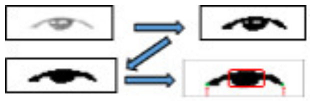

Fig. 8. Steps(binarization, opening and closing to remove noise) in locating iris area.

We locate blink detection by the ratio of the white and black pixel detected mask.When the ratio is bigger than the threshold value, the eyes are considered open and when it is less,they are considered close.

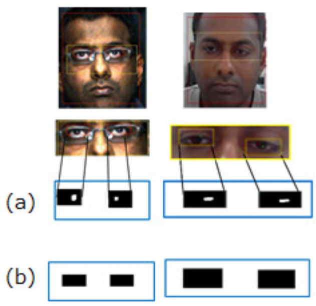

Fig. 9. Detection of frequency of blinking with and without glasses (a)Ratio more than threshold (b) Ratio less than threshold, eyes close.

Facial skin temperature of 3 ROI's (Supraorbital, Periorbital and Maxillary) from thermal domain. In our experiment , measurement is done at three facial areas of sympathetic importance which is periorbital, supraorbital and maxillary in Fig.10. Based on the past researchers [10], [8], [7], we have found out that 3 ROI's are the most affected during mental stress and we focus on the temperature on this areas. At first, we detect the face region using Viola and Jones's boosting algorithm [1], [14] and a set of Cascade structure with Haar-like features. The 3 ROI's are detected based on the face ration Then, the collaboration of ROI with temperature value is done based on the relationship as below. $Temperature = 0.0819 * GrayLevel + 23.762$. Next,we assume the centroid of the detected face area as nose area and nostril mask is used to recalculate the detected area and map into the facial thermal image in Fig.10.

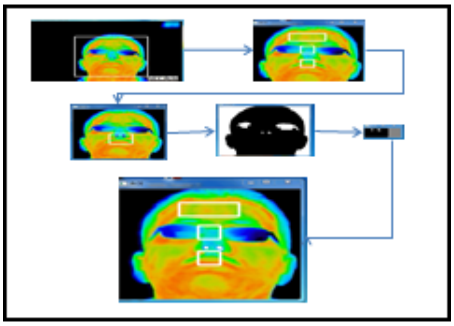

Fig. 10. Proposed automatic Thermal Face, Supraorbital, Periorbital, Maxillary and Nostril Detection.

We then analyze the detected thermal face of the 3 ROI's with sympathetic importance for person wearing spectacles and without it.We found out that there are unwanted ROI such as spectacles and hair which can be excluded so that temperature reading can be done precisely as shown in Fig.11.

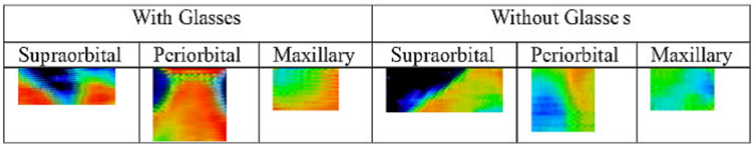

Fig. 11. Detected thermal face with and without spectacles.

To overcome this problem,unwanted ROI is removed by using Graph Cut method as shown in Fig.12.

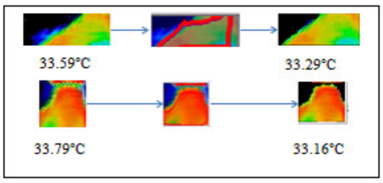

Fig. 12. The unwanted ROI removed using Graph Cut.

Blood vessel volume at supraorbital from thermal domain. Based on recent study conducted, [10], [5], we have found out that user stress is correlated with the increased blood flow in three facial areas of sympathetic importance which is periorbital, supraorbital and maxillary. This increased blood flow dissipates convective heat which can be monitored through mid-wave infrared (MWIR, 3-5μm)camera. Our approach is different which is trying to implement it on a long-wave infrared (LWIR,8-14μm)camera.

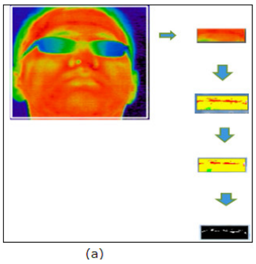

Fig. 13. (a)Segmentation of supraorbital area, white top hat segmentation, bilateral filter and binary image.

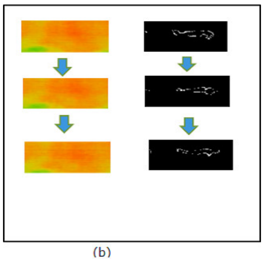

Fig. 14. Detected blood vessel during stress stimulus.

The methodology for detecting blood vessel is shown in the flow chart in Fig.13. After the detection of supraorbital area in thermal IR, image morphing is applied on the diffused image to extract the blood vessels that are relatively low contrast compared with the surrounding tissue. We employ, top hat segmentation method, which is the combination of the erosion and dilation operations. We are interested in the bright(hot) like structure which correspond to blood vessel. For this reason we employ White Top-hat segmentation (WTH) as in the equation below.

$$I \circ S = (I \ominus S) \oplus S, WTH = I - (I \circ S) \tag{3}$$

Where I: is the original image, I$\circ$ S:opened image, S:structuring element,$\ominus$:erosion, $\oplus$: dilation. In order to enhance the edge and reduce noise, bilateral filter is employed. Bilateral filter is a nonlinear, edge preserving and noise reducing smoothing filter. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels.

$$I^{filtered} = \Sigma_{x_i \epsilon \Omega} I(x_i) f_r(||I(x_i) - l(x)||) g_s(||x_i - x||) \tag{4}$$

Where $I^{filtered}$ is the filtered image; I is the original input image to be filtered; x is the coordinates of the pixel to be filtered; Ω is the window centered in x; f_r is the range kernel for smoothing difference in intensities.This function can be Gaussian function, g_s is the spatial kernel for smoothing differences in coordinates. After applying bilateral filter, based on the RGB pixel value, the red range of pixel which is associated to blood vessel is then segmented (Fig. 14). The ground truth comparison images of blood vessel in MWIR thermal IR and Visual, detected in our system are shown in Fig.15.

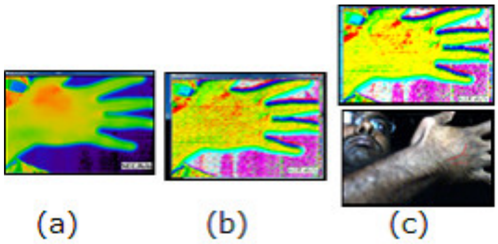

Fig. 15. Hand blood vessel in (a)thermal-actual,(b)thermal-white top hat segmentation (c)thermal-bilateral filter and visual image.

2.3 Experimental setup and result

Experimental Setup Our screening environment consists of light and sound proof screening station, NEC TH7800 thermal camera (right) and USB CMOS Imaging Source DFK 22AUCO3 (left), Stimulus screening monitor,temperature and humidity sensor for monitoring screening station PICO RH-02, Pulse Oximeter SAT-2200 to monitor pulse rate, Nipro cocoro meter to measure salivary amylase level, responder seat and operator machine. About 20 people age from 18-30 both male and female participated in this study. The subject were asked to sit comfortably in the screening station and have him/her rest for about 5-10 minutes before and after the mental stress stimulus test. Mental Stress experiment adopted is designed by Dr Soren Brage of MRC Epidemiology Unit, Cambridge University [2]. It consists of 4 minutes of stroop color word test with 30 second of rest in between and 4 minutes of math test with 30 second of rest in between.

Result of feature matching and detection in Visual and Thermal Camera The result of matching SIFT feature between nostril area in thermal IR and the face part in visual in our pair calibrated Thermal-visible camera is shown in Table 1 below.

Experiment showed 86% of correct matching. To further improve its accuracy, it requires more accurate samples. The development of automatic thermal face, supraorbital, periorbital and maxillary and nostril detection showed 90-96% of correct measurement of ROI and correct temperature was detected. It was calculated based on correctly recognize ROI of 10 correspondent within 4 minutes of stimulus experiment. For detection of eyes blinking in visual domain, almost 98% of eyes blinking were detected successfully for without eye glasses and 89% with glasses.The recognition rate for detected pixel for blood vessel detection was about 93%.

Result of stimulus experiment and emotion recognition system Five experiments were conducted for monitoring before, during and after mental stress stimulus.The first experiment was to measure frequency of eyes blinking in real

Table 1. Results of Matching Features in Thermal-Visible

Frame	Pro. Time[ms]	Fea. Point(TH)	Fea. Point(VI)	Correct Match[%]
1	170.759	2	157	100
2	168.443	2	93	100
3	171.866	11	106	36
4	162.89	2	81	100
5	170.921	2	168	50
6	154.105	1	168	100
7	160.873	1	114	100
8	153.269	1	112	100
9	170.35	1	146	100
10	170.972	1	157	100

time (Fig.16). Before the stimulus, blinking frequency was below 10 times for every second and it doubled to 20 times every second during stimulus and decreased to an average of 13 times afterwards.

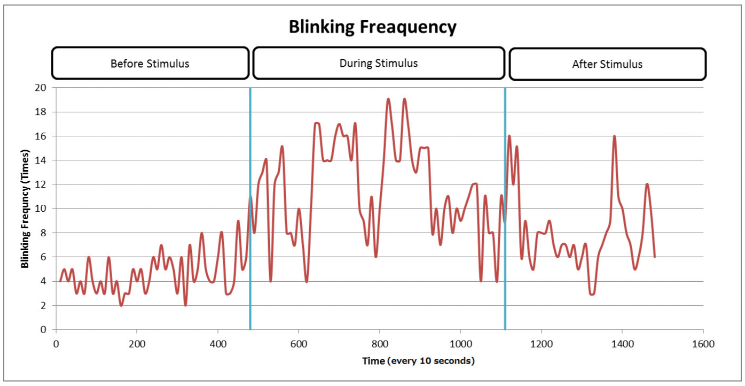

Fig. 16. Frequency of blinking before, during and after stress stimulus .

The second experiment was to monitor the 3 ROI's (Supraorbital, Periorbital and Maxillary) temperatures before, during and after the stress stimulus experiment. In overall, during pre-stress stimulus which was 1-2 min, temperature in the 3 ROI's was stable in value. During stimulus ,which was 2-4 min and 4-6 min, the temperature fluctuated and slight increase tendency can be seen. During rest time which was 4-5 min , there was a slight decrease of temperature at 3 ROI's and it increased gradually after the period. During post-stimulus which was 6-7 min, the temperatures in 3 ROI's were stable and slight temperature decrease tendency can be seen. In most cases, especially in Fig.17.(D) supraorbital area was the most hottest part followed by periorbital and maxillary. In the third

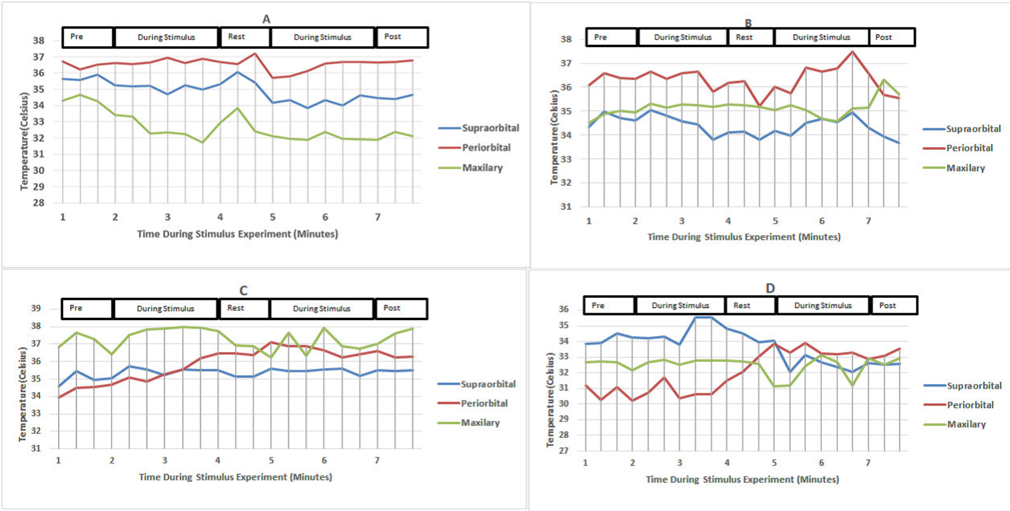

Fig. 17. A-D Temperature of 4 subjects for the first time stimulus experiment conducted at 3 ROI's.

experiment, thermal facial image was also monitored for possible blood vessel which was visible through the stress experiment at the supraorbital area. We have managed to segment out the blood vessel automatically and calculate the pixel value before, during and after experiment in real time. In overall, during rest time which was 4-5 min, less blood vessel pixel was detected in comparison with during mental stress experiment.

In the fourth experiment (ground truth), pulse rate and salivary amylase level was monitored for three correspondent through 2 types of stimulus. In overall, during 10-11 min, there was a slight decrease of pulse rate and salivary amylase level. This physiological measurement of pulse rate and salivary amylase level was consistent and correlate with the other vision based measurement proposed in this research.

In the fifth experiment,classification experiment was done based on Support Vector Machine [11]. Seven features were used as the data entry which were mean value of number of blink frequency, average temperature of 3 ROI, average pixel blood vessel volume, pulse rate and Salivary amylase level. Number of classes were 3 which were before, during and after stimulus experiment. At this stage only 10 experimental subjects were collected and divided into 210 data entries. 20 samples were used as a test samples and the remaining to train the classifier. To evaluate the classification ability of the system under the different conditions, the total classification accuracy, which was the number of correctly classified samples divided by the number of total samples, was calculated for each classification conditions. The overall accuracy reached was 82.46%.

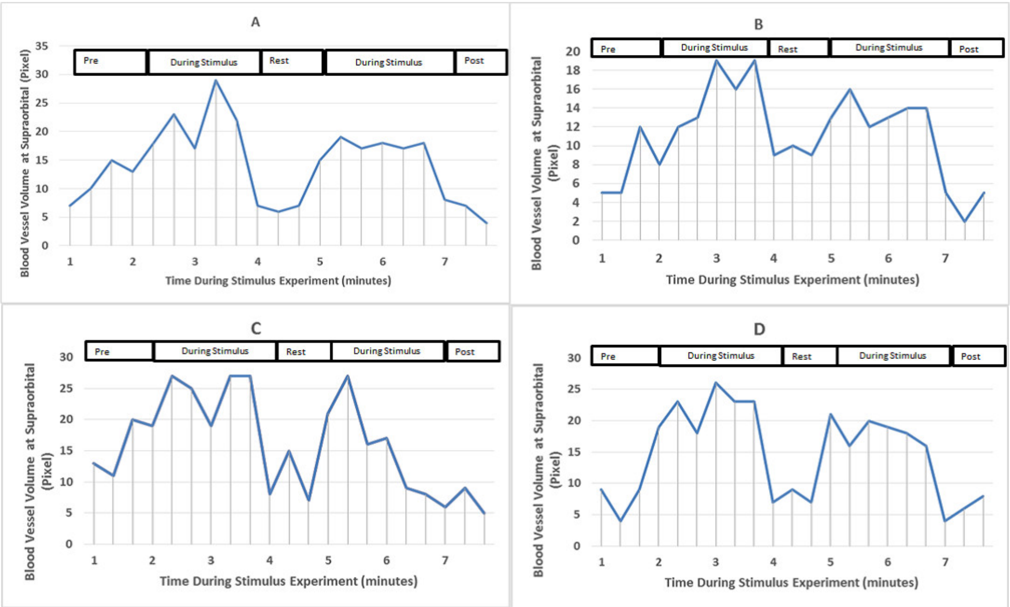

Fig. 18. Blood Vessel Volume at Supraorbital for 4 Correspondent.

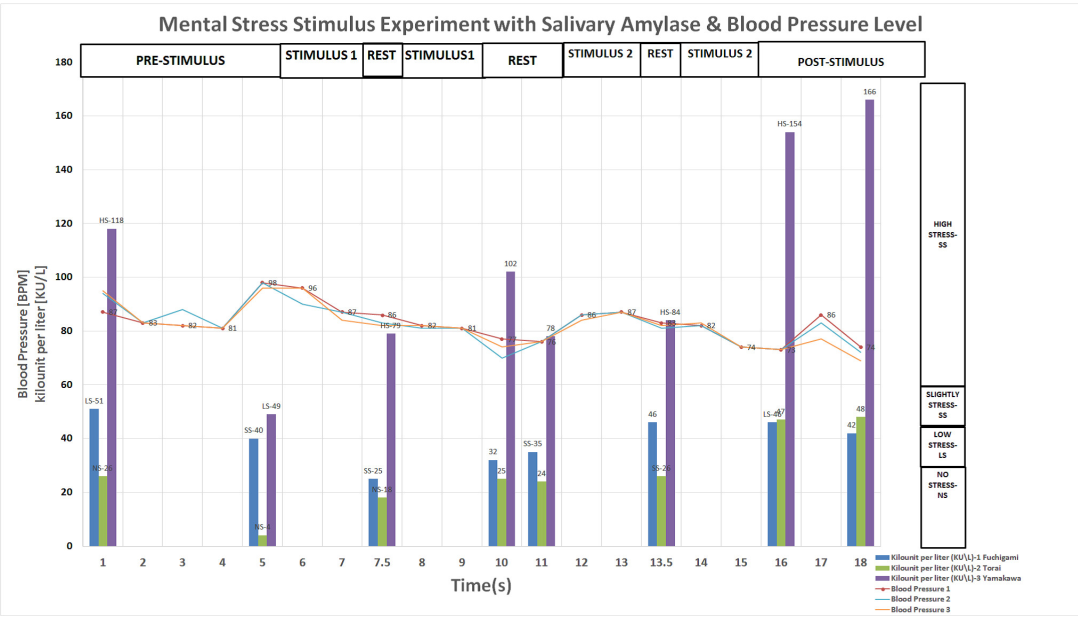

Fig. 19. Blood Pressure and Salivary Amylase level for 3 correspondent.

2.4 Discussion

In this research, we only consider frontal face in both thermal and infrared case as the stimulus monitor is located in front of the correspondent in our screening environment. However, facial occlusion in stereo visual-infrared is still a challenging issue. In our first experiment, the recognition rate of eyes blinking in glasses case is slightly low from without glasses because of the shadow from the light which appears on the glass surface and the frame of the glasses. However it can be overcome by implementing proper screening environment. From this real time data it shows that blinking frequency is highly correlated with mental stress. More experiment data should be taken in the later stage. In the second experiment, we have also found out that user stress is correlated with the increased blood flow in three facial areas of sympathetic importance which is periorbital, supraorbital and maxillary which somehow increase the temperature during stimulus. From the third and fourth experiment, we can conclude that mental stress is highly correlated with the activation of the corrugator muscle on the forehead. However segmenting the thermal imprints of the supraorbital vessels is challenging because they are fuzzy due to thermal diffusion and exhibit significant inter-individual and intra-individual variation. On the average the diameter of the blood vessels is 10-15m, which is too small for accurate detection and 0.1°C warmer than the adjacent skin. In the fifth experiment, the overall SVM classification accuracy is considered high at 82.46%. C.D.Katsis et. al [12] reported accuracy of only 79.3% using SVM classification method by features extracted from EMG, ECG, respiration and EDA biosignals. Moreover, the image processing techniques is not used by [12] to extract facial characteristic as the the drivers need to wear a helmet. Physiological measurement of pulse rate and salivary amylase level are considered to be the most reliable measurement of mental stress according to a lot of researchers and they are used as a ground truth value in our research. They also shows correlation with other vision based measurement proposed in this research.

2.5 Conclusion

Our objective is to propose vision-based measurement as an alternative to physiological measurement. In this research three vision-based and two physiological measurement had been proposed and shows promising results. Usually in physiological measurement, combination of measures may be redundant with others and this may cause collection of unnecessarily large volumes of data and unnecessary processing time. This motivates the use of vision-based as it only requires crucial data after pattern recognition and processing of partial image. Our methodology to estimates emotional state from human subjects by extracting facial characteristic shows good performance.

Acknowledgement. This work was partially funded by MEXT/JSPS Kakenhi grant number: 50325768 and University Tun Hussein Onn Malaysia (UTHM). We would like to give special thanks to the laboratory members for their invaluable inputs and assistance.

References

1. Viola,P.A.,Jones and M.J.: Rapid object detection using a boosted cascade of simple features. IEEE,CVPR,(2001) 511-518
2. N.Sharma, T. Gedeon: Objective measures, sensors and computational techniques for stress recognition and classification: A survey. Elsevier, Computer Method and Program in Biomedicine **3** (2012) 1287-301
3. Noriyasu Takai, Masaki Yamaguchi et. al.: Effect of physiology stress on the salivary cortisol and amylase levels in healthy young adults. Archives of Oral Biology, Elsevier **49** (2004) 963-968
4. M.Akhloufi, A. Bendada, Jean Christphone Batsale: State of the art in infrared face recognition. QIRT Journal (2008) 1-24
5. J.Levine, I.Pavlidis, M.Cooper: The face of fear. Lancet **357** (2001) 1757
6. Hartley R. and Zisserman A.: *Multiple view geometry in computer vision.* Cambridge University Press(2000)
7. Eberhard Tan, J.Levine: Human Behavior: Seeing through the face of deception. Nature **415** (2002) 6867
8. Dvijesh A.Merla, P. Tsiamyrtzis,I.Pavlidis: Imaging Facial Signs of Neurophysiological Responses. IEEE Transactions on Biomedical Engineering **56** (2009) 2
9. David G. Lowe: Distictive image features fromscale-invariant keypoints. International Journal of computer vision (2004) 60
10. D.Shatri M. Papadakis and P. Tsiamyrtzis: Perinasal imaging of physiological stress and its affective potential. IEEE Transaction Affective Computing (2012) 366-378
11. Chang,C.C.Lin, C.J: LIBSVM:a library for support vector machines. http://www.csie.ntu.edu.tw/ cjlin/libsvm (2001)
12. C.D.Katsis , Nikolas Katertsidis et. al.: Toward emotion recognition in car-racing dirvers: a biosignal processing approach. IEEE Transactions on System, Man and Cybernetics, Part A: System and Humans **38** (2008) 502-512
13. A.Majumder, L. Bahera and V.K. Subramanian: Automatic and Robust Detection of facial Features in Frontal Face Images. IEEE Transactions on System, Man and UKsim Int. Conference on Computer Model and Simulation **3** (2011) 331-336
14. A.Majumder, L. Bahera and V.K. Subramanian: Automatic and Robust Detection of facial Features in Frontal Face Images. IEEE Transactions on System, Man and UKsim Int. Conference on Computer Model and Simulation **3** (2011) 331-336
15. Mohd Norzali Haji Mohd, Masayuki Kashima, Kiminori Sato and Mutsumi Watanabe: Thermal-visual facial feature extraction based on nostril mask. IAPR-Machine Vision Application **4-20** (2013) 113-116
16. Mohd Norzali Haji Mohd, Masayuki Kashima, Kiminori Sato and Mutsumi Watanabe: Effective Geometric Calibration and Facial Feature Extraction Using Multi Sensors. International Journal of Engineering Science and Innovative Technology (IJESIT) **1-2** (2012) 170-178
17. Jiang-Gang Wang: Facial Feature extraction in an infrared image by proxy with visible face image. IEEE Transactions on Affective Computing **3-3** (2007) 2057-2066