Video-Based Contactless Assessment of Physiological Signals and Facial Features

[1] J.Dayalin Subi,[2]I.Gracy Joylet,[3]X.Axlin

[1] dayalinsubi@gmail.com,[2]gracyjoylet11@gmail.com,[3]axlinxavier4@gmail.com

*[1,2,3]Assistant Professor, Dept. of Biomedical Engineering,
Loyola Institute of Technology and Science, Thovalai*

Abstract – We explored how computer vision techniques can be used to detect engagement while students (N = 22) completed a structured writing activity (draft-feedback-review) similar to activities encountered in educational settings. Students provided engagement annotations both concurrently during the writing activity and retrospectively from videos of their faces after the activity. We used computer vision techniques to extract three sets of features from videos, heart rate, Animation Units (from Microsoft Kinect Face Tracker), and local binary patterns in three orthogonal planes (LBP-TOP). These features were used in supervised learning for detection of concurrent and retrospective self-reported engagement. Area Under the ROC Curve (AUC) was used to evaluate classifier accuracy using leave-several-students-out cross validation. We achieved an AUC = .758 for concurrent annotations and AUC = .733 for retrospective annotations. The Kinect Face Tracker features produced the best results among the individual channels, but the overall best results were found using a fusion of channels.

Keywords **–** Image Preprocessing**,** Feature extraction**,** Detection

I. INTRODUCTION

Learning-centered affective states, such as engagement, are inextricably linked with the cognitive aspects of learning. It is widely acknowledged that the way users engage in an activity is an essential component of their experience with the activity. The way people engage with an activity has been studied from multiple perspectives in HCI and psychology. The term engagement itself is interpreted in somewhat different ways by different com-munities of researchers, but most definitions maintain that engagement involves attention and emotional involvement with a task. Engagement is also not stable, but fluctuates throughout an interaction experience. In the area of HCI, discuss four phases of engagement: the beginning (i.e. point) of engagement, sustained attention or engagement, disengagement and reengagement. Our present emphasis is on engagement during learning (or educational activities). Many authors (c.f.) agree on four types of engagement during learning. Behavioral engagement can be assessed by observing persistence and effort; emotional engagement can be assessed by detecting supportive emotions (e.g., interest, curiosity) and self-efficacy. Cognitive engagement is demonstrated when the student shows a sophisticated approach to the activity, for example by using deep rather than

Measuring engagement has been important in educational research because it allows researchers to under-stand what decisions promote or hinder engagement. Studies that focus on students ‘engagement need a way of measuring it. This can be done with one of the two types of data identified by engagement theorists: internal to the individual (cognitive and affective) and external observable factors, such as perceptible facial features, postures, speech, and actions. Methodologically, as is common in many affective computing applications (c.f.), studying engagement requires bringing together observational data (e.g., facial expressions, speech) and subjective data (e.g., self-reports). New sensing and affective computing techniques allow for novel methodological approaches to measuring engagement. Different modalities such as video, audio and physiological measures are being used for affect detection in learning contexts. Multimodal approaches have also been explored to improve the accuracy of affect detection in learning applications.

The emotional states of students can be inferred from these measures via affective computing techniques that are increasingly being used in learning technologies. In this project, we attempt to detect engagement in a different educational task (writing) with a rather different methodological approach. This work focus on writing is motivated by the fact that writing is one of the most common activities in both work and educational contexts, so we aim to support writing tools that help students engage with and enjoy their writing activities.

Image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image.

Digital image processing is the use of computer algorithms to perform image processing on digital images.

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable(quantitative variable) and was first introduced by Karl Pearson.

A.Facial Expression

A facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication. They are a primary means of conveying social information between humans, but they also occur in most other mammals and some other animal species.

Beyond the accessory nature of facial expressions in spoken communication between people, they play a significant role in communication with sign language. Many phrases in sign language include facial expressions in the display. There is controversy surrounding the question of whether or not facial expressions are worldwide and universal displays among humans. Supporters of the Universality Hypothesis claim that many facial expressions are innate and have roots in evolutionary ancestors. Opponents of this view question the accuracy of the studies used to test this claim and instead believe that facial expressions are conditioned and that people view and understand facial expressions in large part from the social situations around them.

B.Heart Rate

Heart rate is the speed of the heartbeat measured by the number of contractions of the heart per minute (bpm). The heart rate can vary according to the body's physical needs, including the need to absorb oxygen and excrete carbon dioxide. It is usually equal or close to the pulse measured at any peripheral point. Activities that can provoke change
include physicalexercise, sleep, anxiety, stress, illness, and ingestion of drugs.

Many texts cite the normal resting adult human heart rate range from 60–100 bpm. Tachycardia is a fast heart rate, defined as above 100 bpm at rest. Bradycardia is a slow heart rate, defined as below 60 bpm at rest. Several studies, as well as expert consensus indicate that the normal resting adult heart rate is probably closer to a range between 50-90 bpm. During sleep a slow heartbeat with rates around 40–50 bpm is common and is considered normal. When the heart is not beating in a regular pattern, this is referred to as an arrhythmia. Abnormalities of heart rate sometimes indicate disease.

II. RELATED WORK

The paper titled “Automatically Recognizing Facial Expression: Predicting Engagement and Frustration,” published by Joseph F. Grafsgaard, Joseph B. Wiggins. In this paper, an automated analysis of fine-grained facial movements that occur during computer-mediated tutoring is presented. Here use the Computer Expression Recognition Toolbox (CERT) to track fine-grained facial movements consisting of eyebrow raising (inner and outer), brow lowering, eyelid tightening, and mouth dimpling within a naturalistic video corpus of tutorial dialogue (*N*=65). Within the dataset, upper face movements were found to be predictive of engagement, frustration, and learning, while mouth dimpling was a positive predictor of learning and self-reported performance. These results highlight how both intensity and frequency of facial expressions predict tutoring outcomes. Additionally, this paper presents a novel validation of an automated tracking tool on a naturalistic tutoring dataset, comparing CERT results with manual annotations across a prior video corpus. Disadvantages of this method is high computational cost.

The paper titled **“**Facial Action Recognition for Facial Expression Analysis from Static Face Images”published by MajaPantic, and Leon J. M. Rothkrantz. In this paper, an automated system is presented. Tha.developed to recognize facial gestures in static, frontal- and/or profile-view color face images. A multi detector approach to facial feature localization is utilized to spatially sample the profile contour and the contours of the facial components such as the eyes and the mouth. From the extracted contours of the facial features, we extract ten profile-contour fiducial points and 19 fiducial points of the contours of the facial components. Based on these, 32 individual facial muscle actions (AUs) occurring alone or in combination are recognized using rule-based reasoning. With each scored AU, the utilized algorithm associates a factor denoting the certainty with which the pertinent AU has been scored. Disadvantage of this method is low detection accuracy.

The paper titled “Fully Automatic Facial Action Recognition in Spontaneous Behavior**”** published by Anonymous. In this paper presented a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units. We present preliminary results on a task of facial action detection in spontaneous expressions during discourse. Support vector machines and AdaBoost classifiers are compared. For both classifiers, the output margin contains information about action unit intensity. Disadvantage of this method is cless effective.

The paper titled “Dynamics of facial expression extracted automatically from video” published by Gwen Littlewort , Marian Stewart Bartlett, Ian Fasel. In this paper presented a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions, including AdaBoost, support vector machines, and linear discriminant analysis. Each video-frame is first scanned in real-time to detect approximately upright-frontal faces. The faces found are scaled into image patches of equal size, convolved with a bank of Gabor energy filters, and then passed to a recognition engine that codes facial expressions into 7 dimensions in real time: neutral, anger, disgust, fear, joy, sadness, surprise. Disadvantage of this method is high complexity.

The paper titled "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions" published by Jacob Whitehill, ZewelanjiSerpell, Yi-Ching Lin. In this paper presented an approaches for automatic recognition of engagement from students' facial expressions. We studied whether human observers can reliably judge engagement from the face; analyzed the signals observers use to make these judgments; and automated the process using machine learning. We found that human observers reliably agree when discriminating low versus high degrees of engagement. When fine discrimination is required (four distinct levels) the reliability decreases, but is still quite high. Furthermore, we found that engagement labels of 10-second video clips can be reliably predicted from the average labels of their constituent frames, suggesting that static expressions contain the bulk of the information used by observers. We used machine learning to develop automatic engagement detectors and found that for binary classification (e.g., high engagement versus low engagement), automated engagement detectors perform with comparable accuracy to humans. Disadvantage of this method is only detecting the facial expression.

III. METHODOLOGY

Most previous work of affect detection has focused on detecting basic emotions, but more recently some researchers have focused on the recognition of complex mental states, particularly attention and engagement. Engagement can be measured from different behavioral expressions: eye-gaze movements, facial features, gestures, and so on. Unfortunately, eye tracking is affected by head movements and is not yet easily scalable in real-world contexts. The present emphasis is on physiology-based and facial-feature based engagement detection as these are the two methods we explore in this research.

A.Physiology-based detection

Central and peripheral physiological signals have been commonly used for detecting task engagement, alertness, and drowsiness. Most of the proposed methods for measuring physiological states attempt to record and analyze the electrical signals produced by heart, brain, muscles and skin. The main instruments used to monitor physiological signals include Electrocardiogram (ECG), Electromyogram (EMG), galvanic skin response (GSR), and Respiration (RSP). Electroencephalogram (EEG) is widely used to differentiate between alertness vs. drowsiness. Various EEG-based engagement indices have been proposed. Classification accuracies between 84% and 99% have been achieved for detecting driver drowsiness detection using EEG methods.

Cardiac activity has also been explored for automatic affect and engagement/alertness detection. Heart rate (HR) and heart rate variability (HRV) are two important ECG measures which have been used widely for these purposes. Previous researches showed that HR is a good indicator for discriminating between different affective states. For example, the HR tends to be higher during fear, anger and sadness than during happiness, disgust and surprise. HR and HRV have been shown to be indicators of alertness and drowsiness.

One of the main challenges associated with physiological based affective computing applications is the intrusiveness of physiological sensors. Users must have access to a heart rate monitor, which typically must be physically attached to the skin. This issue can be addressed by using remote measurement techniques. Three different approaches have been investigated for remote, contactless measurement of vital signs such as heart rate. Microwave Doppler radar was one of the earliest methods for sensing heart rate and respiration. Thermal imaging is another approach for heart rate detection using analysis of skin temperature modulation. More recent approaches include video-based imaging methods that use photoplethysmography to detect HRV.

B.Facial-feature based detection

Two main approaches have typically been used in the area of facial expression analysis: geometric-based and appearance-based approaches. Geometric features include shapes and positions of face components, and the location of fixed facial points such as the corners of the eyes, eyebrows. Appearance-based methods recognize facial expressions by analyzing the changes of the face‘s surface in both static and dynamic space (e.g., dynamic texture-based techniques). Facial expression recognition systems that use appearance-based features have been reported. Several re-searchers have used different types of features: for example, Gabor wavelet coefficients, optical flow, and Active Appearance Models.

There are strengths and weaknesses in both the geometric based and appearance based approaches. Geometric based methods typically track the position of a number of facial points in time. With this approach, some features of facial appearances (e.g., shape of mouth, position of eye-brows) can be extracted, while features related to texture of the face (e.g., furrows and wrinkles) cannot be extracted. In contrast, appearance based methods may be more sensitive to changes in illumination (e.g., brightness and shadows), head motions and differences between shapes of the faces. They claimed that the geometric features outperformed the appearance features, yet using a combination of both yielded the best results.

Affect and engagement detection from facial features has also been investigated in learning contexts. For example,used the Computer Expression Recognition Toolbox (CERT) to track facial movements within a naturalistic video corpus of tutorial dialogue. The most frequent AUs including eyebrow raising (inner and outer), brow lowering, eyelid tightening, and mouth dimpling were selected to predict overall levels of engagement, frustration and learning gains using forward stepwise linear regression. Their findings suggested that upper face movements would be a reliable predictor of engagement, frustration, and learning. They achieved reasonable agreement between their predictions and manual annotations, albeit at a rather coarse grained level (i.e., across the entire learning session).

C.Proposed system

Improved automatic detection of engagement in computerized education environments will lead to more effective learning and a more engaging experience for students. In this project a new method is introduced to detect engagement in an activity. A combination of geometric features (STFT), appearance features (Local binary patterns in three orthogonal planes), and physiological features (heart rate) that were extracted using computer vision techniques.

The proposed approach emphasis is on both physiology-based and facial-feature based engagement detection methods are used to detect the engagement. In the first step, three types of features were extracted from video. After feature extraction, facial features, heart rate features and local binary pattern features are extracted from the input video. Finally, Support Vector Machine (SVM) classifier is used to detect whether the person is engaged in an activity or not from the extracted features.

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be very redundant, then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Geometrical features are extracted using Short Time Fourier Transform (STFT). Appearance features are extracted using Local Binary Pattern (LBP). Finally, Heart rate features are extracted from the face. Support Vector Machine (SVM) classifier is used to detect whether the person is engaged in an activity or not from the extracted features.

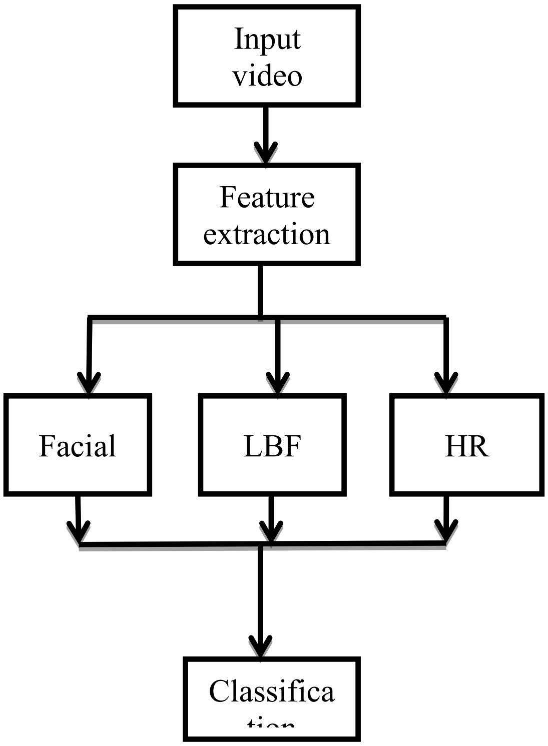


Fig.1 proposed system

In this project three main steps for detecting student’s-gagement levels are followed. In the first step, three types of features (Facial features, LBP, and Heart Rate - HR) were extracted from each video segment. The last 10 seconds of video before each annotation was considered for concurrent segments. The features were synchronized with corresponding concurrent and retrospective labels. Finally, machine learning classification techniques were applied on features and validated with leave-several-students-out cross-validation for student-independent models.

Transforming the input data into the set of features is called feature extraction. A two-dimensional function of time and frequency of a signal can be obtained by short time Fourier transform (STFT). An expedient acceptance between the time and frequency based views of a signal can be represented by STFT. It gives some explanation about both when and at what frequencies a signal event occurs.

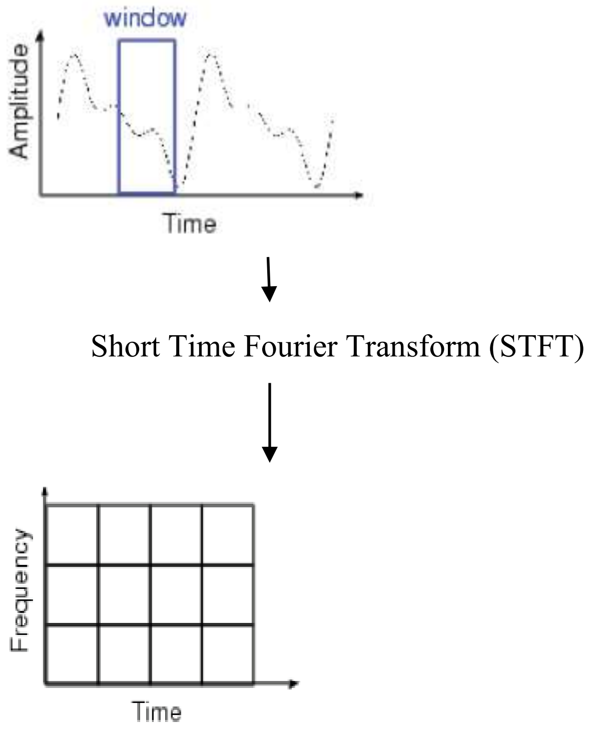


Fig.2 Short time Fourier transforms

Heart rate is the speed of the heartbeat measured by the number of contractions of the heart per minute (bpm).The frame rate is set at 30 frames per second (fps). The resolution of the video frame is the QQVGA, which are 160 by 120 pixels.

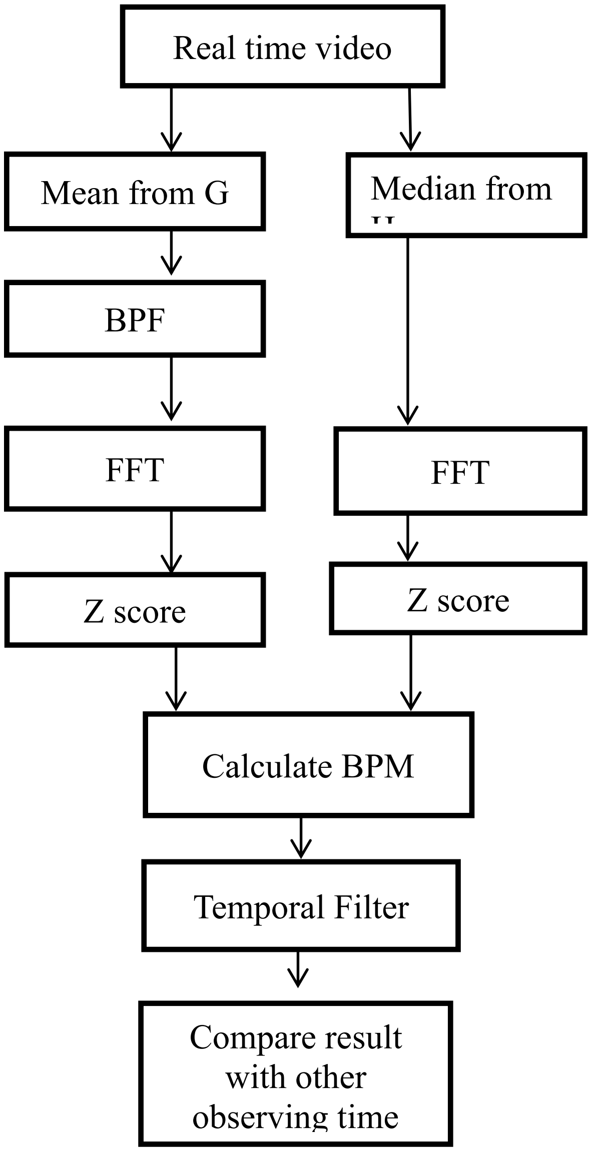


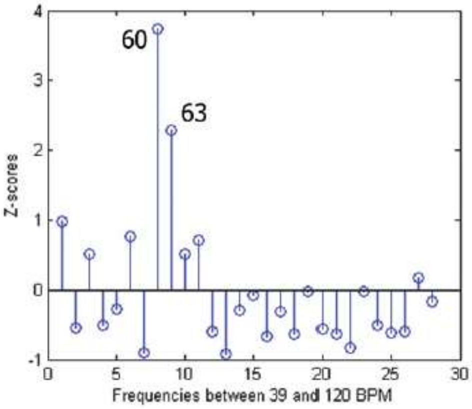
Fig.3 Block diagram of proposed method

The sample frame of an input video sequence during the experiment. The rectangular box shows a region of interest (ROI), which is manually selected. The participants are asked to stay still in order that their faces are present within the ROI box during the measurement.

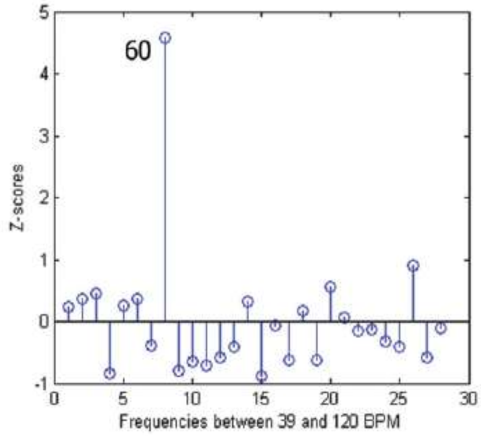
Fig.4 samples frame of an input video sequence for heart rate measurement

They are also asked to hold a smartphone to measure their heart rates (HR) using the application Cardiio and use it as a ground truth. The proposed methods update HR measurements every two seconds. Fig. 4.4 shows a flow chart for HR features extraction in each observing time. During the experiment, the G and H signal are use in the method. It has been said that the H signal in human skin color are all the same color, for example, black and whites skin. The G signal extract from a video sequence directly while H signal is computed from R, G, and B signals.

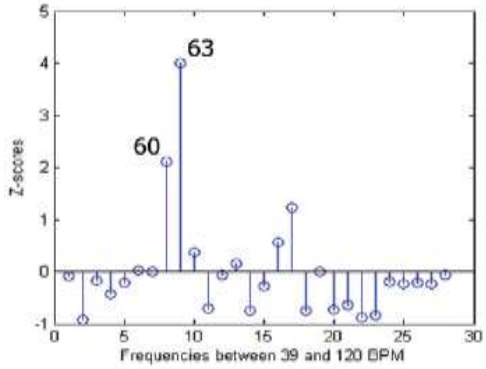
Then, calculate the mean value from G signal and median value for H signal and keep this in different amount of recent data for multi-observing time purpose which is 10 and 20 seconds. In each observing times, we apply a band-pass filter (BPF) with the filter mask [-1, 0, 1] to the G signal data set and then apply the fast Fourier transform (FFT) to the two signals.



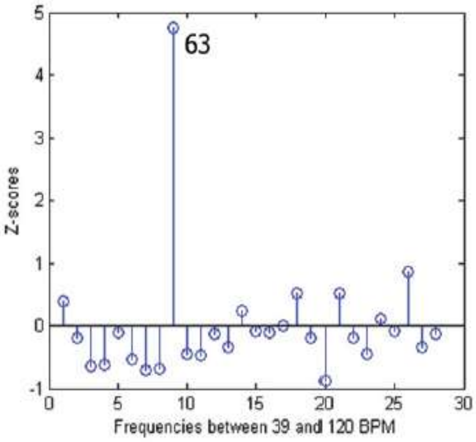
(a)



(b)



(c)



(d)

Fig. 5 Z-scores of the spectrums between 39 and 120 BPM. Input signals are sinusoidal curves of difference frequencies (a) 60, (b) 61, (c) 62, and (d) 63 BPM with random noise

We set the lower and upper limits of the HR from 40 and 120 beats per minute (BPM) that is frequency range between 0.67 and 2.0Hz. The Fourier spectrums within this frequency interval are used for evaluated by z-scores expressed as,

$$Z = \frac{X-m}{\sigma_x}$$

Where, x is the Fourier spectrums. m and σ_x mean and standard deviation of x in the selected band.. We first find the spectrums with the first and second largest z-scores. if the two most prominent spectrums are next to each other, for instance, the nth spectrum with the largest z-score and the $(n+1)^{th}$with the second largest, we calculate the BPM using both of them.

When the observing time is 20 seconds, the n^{th} spectrum corresponds to 3n BPM, while $(n+1)^{th}$ results in 3(n+1) BPM. Fig. 4.6(b) demonstrates that there is the most distinct spectrum at the frequency that corresponds to 60 BPM. The second most prominent spectrum indicates 63 BPM. There are two choices, 61 and 62 BPM, between them. We then select 61 BPM as the final result because it makes more sense to place a higher priority on the most prominent spectrum than on others.

Similarly, when the most distinct spectrum indicates 63 BPM and the 2nd points 60 BPM as shown in Fig. 4.6(c), the final selection will be 62 BPM. If the two most prominent spectrums are not next to each other, as exhibited in Figs. 4.6(a) and (d), simply use the single spectrum with the largest z-score. In this way, coarse estimation of the HR because of the limited observing time (such as 20 seconds) can be refined. The measurements by the G and H signals are compared within the same observing time. The measurement with a higher z-score is selected as a winner. Furthermore, a temporal filter is applied to the 5 most recent measurements where the measurement with the largest z-score among them is selected as output.

HR measurement may occasionally include faulty BPMs because of external factors, for instance, varying illuminations and the motions of the face. It is possible to reject those faulty BPMs by monitoring the measurements over time (though a short period) and select one with a high confidence, i.e. z-score. This is how our temporal filter works. The filter is particularly effective for improving the accuracy of the HR measurements in a short observing time. In this way, obtain 2 HR measurements from two different observing times.
Fig.4.7 shows how to determine the final HR from the two measurements. Here compare the measurements from 10 second and 20 second observations in the stage of Selection. If two measurements are similar to each other, the result from 20-second observation is selected as a winner. If the two measurements are significantly different from each other, the result from 10-second observation is selected. This is because the measurement with a longer observation time tends to be more accurate when the HR is stable.

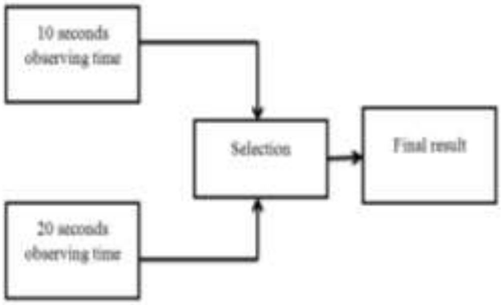


Fig.6 The block diagram of HR measurements from 2 different observing times

The measurement with a shorter observation is more effective when the HR varies because it responds to the change of HR more quickly. In this manner, we attempt to achieve both fast response and high accuracy in HR measurement. The proposed method is fast enough to perform HR measurement in real time and it can update the measurement every two seconds. The proposed method is not limited to the measurement times 10 and 20 seconds. A point is that proposed method is designed for obtaining the advantages of short and long observing times.

A fastFourier ransform algorithm computes the discrete Fourier transform(DFT) of a sequence, or it's inverse. Filtering in time and/or space is a long-established method in any signal detection process to help "clean up" your signal. Temporal filtering aims to remove or attenuate frequencies within the raw signal that are not of interest.

The proposed system is implemented using MATLAB. Here, a new method is proposed to detect engagement in an activity. Then the following steps are done in this project. First we had loaded the input video. Hereinafter preprocessed the input image. The preprocessing steps are includes frame conversion and face detection. First the input video is converted in to frame. In feature extraction, three features are extracted from the input video. Short time Fourier transforms (STFT) is used to extract the facial features. Local Binary Pattern (LBP) is used to extract the appearance features. Finally, Support Vector Machine (SVM) classifier is used to detect whether the person is engaged in an activity or not from the extracted features. Support vector machine are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis Support vector machine are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

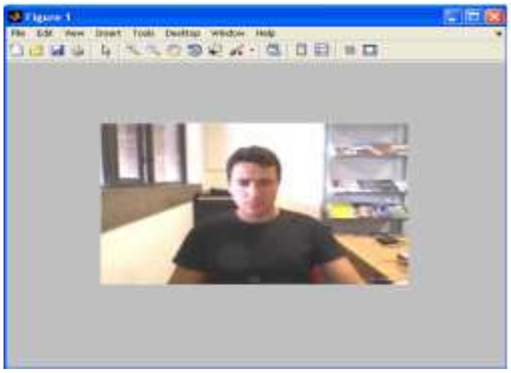


Fig .7.1 Input video

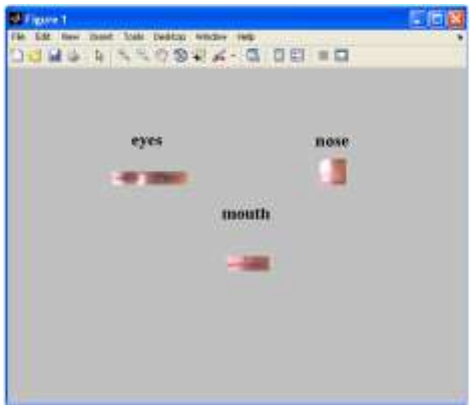


Fig.7.2 Extracted facial features

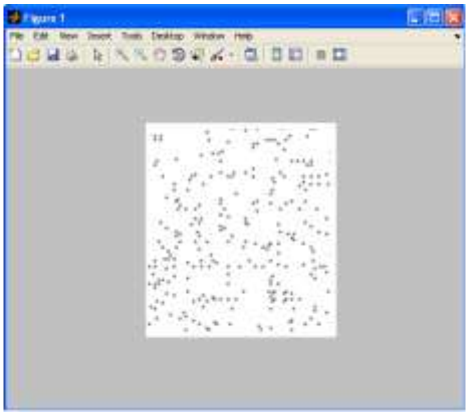


Fig .7.3 Extracted LBP features

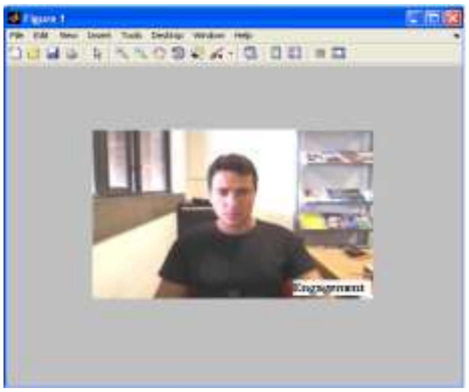


Fig .7.4 Classified image

IV. CONCLUSION

Learning involves a rich array of cognitive and affective states. Recognizing and understanding these cognitive and affective dimensions of learning is key to designing informed interventions. Prior research has highlighted the importance of facial expressions in learning-centered affective states, but tracking facial expression poses significant challenges. It is our hope that improved automatic detection of engagement in computerized education environments will lead to more effective learning and a more engaging experience for students. To that end we presented methods and results for detecting student engagement in the context of a writing task. These methods showed that combining facial texture- and appearance- based features resulted in the most accurate student independent engagement detectors. Then the HR features are extracted from the face. SVM classifier is used to classify the engagement in an activity. The proposed engagement detection methods are very effective and improve the classification accuracy.

REFERENCES

1. C. Peters, G. Castellano, and S. de Freitas, ―An exploration of user engagement in HCI in Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots - AFFINE '09, 2009, pp. 1–3.

2. S. L. Christenson, A. L. Reschly, and C. Wylie, Eds., Handbook of Research in Student Engagement. New York: Springer, 2012.

3. E. A. Linnenbrink and P. R. Pintrich, ―The Role of Self-Efficacy Beliefs Instudent Engagement and Learning Intheclassroom, Read. Writ. Q. Overcoming Learn. Difficulties, vol. 19, no. 2, pp. 119–137, Apr. 2003.

4. J. Reeve and C.-M. Tseng, ―Agency as a fourth aspect of students' engagement during learning activities,Contemp. Educ. Psychol., vol. 36, no. 4, pp. 257–267, Oct. 2011.

5. W. A. Kahn, ―Psychological Conditions of Personal Engagement and Disengagement at Work Acad. Manag. J., vol. 33, no. 4, pp. 692–724, 1990.

6. R. A. Calvo and D. Peters, Positive Computing: Technology for Wellbeing and Human Potential. Cambridge, MA: MIT Press, 2014.

7. R. A. Calvo, S. K. D'Mello, A. Kappas, and J. Gratch, Eds., The Oxford Handbook of Affective Computing. New York: Oxford University Press, 2014.

8. J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, J. C. Lester, and N. Carolina, ―Embodied Affect in Tutorial Dialogue : Student Gesture and Posture,in Proceedings of the 16th international conference on Artificial intelligence in education, 2013, pp. 1–10.

9. K. Forbes-riley and D. Litman, ―When Does Disengagement Correlate with

Learning in Spoken Dialog Computer Tutoring in Proceedings of the 15th international conference on Artificial intelligence in education, 2011, pp. 81–89.

10. M. S. Hussain, O. Alzoubi, R. A. Calvo, and S. D. Mello, ―Affect Detection from Multichannel Physiology during Learning Sessions with AutoTutor in Artificial Intelligence in Education, G. Biswas, S. Bull, J. Kay, and A. Mitrovic, Eds. Springer, 2011, pp. 131–138.

11. A. Kapoor and R. W. Picard, ―Multimodal affect recognition in learning environments,in Proceedings of the 13th annual ACM international conference on Multimedia, 2005, pp. 677–682.

12. J. Whitehill, Z. Serpell, A. Foster, Y.-C. Lin, B. Pearson, M. Bartlett, and J. Movellan, ―Towards an Optimal Affect-Sensitive Instructional System of cognitive skills in IEEE Conference on Computer Vision and Pattern Recognition: Workshop on Human-Communicative Behavior, 2011, pp. 20–25.

13. A. Graesser, A. Witherspoon, B. McDaniel, S. D'Mello, P. Chipman, and B. Gholson, ―Detection of Emotions during Learning with AutoTutor,in Proceedings of the 28th Annual Meetings of the Cognitive Science Society, 2006, pp. 285–290.

14. R. A. Calvo and S. D'Mello, Eds., New Perspectives on Affect and Learning Technologies, vol. 3. New York: Springer, 2011.

15. M. Soleymani and M. Larson, ―Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus in Proceedings of the ACM SIGIR 2010 workshop on crowdsourcing for search evaluation (CSE 2010), 2010, pp. 4–8.

16. J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, ―The Faces of Engagement: Automatic Recognition of Student Engagement from Facial ExpressionsIEEE Trans. Affect. Comput., vol. 5, no. 1, pp. 86–98, Jan. 2014.

17. S. Afzal and P. Robinson, Natural Affect Data - Collection & Annotation in a Learning Context,in3rd International conference on Affective Computing and Intelligent Interaction, ACII 2009, 2009, pp. 1–7.

18. S. D'Mello and R. A. Calvo, ―Beyond the Basic Emotions: What Should Affective Computing Computein CHI 2013 - Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013.

19. Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, ―A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.