

MINISTRY OF EDUCATION AND TRAINING
LAC HONG UNIVERSITY



TRAN THANH PHUONG

**SOME FACIAL EXPRESSION RECOGNITION TECHNIQUES
FOR ASSESSING LEARNER ENGAGEMENT**

COMPUTER SCIENCE PHD THESIS ABSTRACT

Major: Computer Science

Code: 9480101

SCIENCE INSTRUCTOR

Assoc. Prof. Dr. DO NANG TOAN

PhD. LAM THANH HIEN

Dong Nai, year 2023

The work is completed at: Lac Hong University

Science instructor:

1. Assoc. Prof. Dr. Do Nang Toan
2. PhD. Lam Thanh Hien

Review 1:

Review 2:

Review 3:

The thesis will be defended in front of the institutional examination council

.....
.....
.....

At hour day month year

The thesis can be found at the library:

- The library of Lac Hong University
- National Library

PUBLISHED WORKS

- [CT1]. **Tran Thanh Phuong**, Lam Thanh Hien, Ngo Duc Vinh, Ha Manh Toan, Do Nang Toan (2023), “An algorithm for decomposing variations of a 3D model”, *International Journal of Electrical and Computer Engineering (IJECE)*.
- [CT2]. **Tran Thanh Phuong**, Ngo Duc Vinh, Ha Manh Toan, Nong Minh Ngoc, Do Nang Toan (2023), “Simulating the dynamic movements of a 3D human head model based on Vietnamese speech”, *TNU Journal of Science and Technology*, Vol 228, No.15, tr20-28.
- [CT3]. **Tran Thanh Phuong**, Lam Thanh Hien, Do Nang Toan, Ngo Duc Vinh (2022), “An Eye Blink detection technique in video surveillance based on Eye Aspect Ratio”, *Proceedings of International Conference on Advanced Communication Technology (ICACT)*, 13-16 Feb. 2022, Pyeong Chang, Korea, pp534-538
- [CT4]. **Tran Thanh Phuong**, Tran Van Lang, Do Nang Toan (2021), “A technique of assessing student engagement based on eye close/open behavior”, *TNU Journal of Science and Technology*, Vol 226, No.07, tr262-269.
- [CT5]. **Tran Thanh Phuong**, Lam Thanh Hien, Do Nang Toan, Ngo Duc Vinh (2021), “A technique for detecting the open/closed state of eyes based on ratio characteristics”, *Proceedings of the National Conference "Selected Issues in Information Technology and Telecommunications."*, Thai Nguyen 13-14/12/2021, tr265-269.

ABSTRACT

Automatic learning analysis is becoming a significant topic in the education community, where efficient systems are needed to monitor the learning process and provide feedback to teachers. Recent advances in visual sensors and computer vision methods allow for the automatic tracking of behavior and emotional states of learners at various levels. Emotional states such as interest, fatigue, and confusion are automatically identified from facial expressions, and attention states are computed from various visual cues such as looking at faces and body posture. Understanding the behavior of learners helps schools improve training programs, learning environments, and equipment, among other aspects. Additionally, it assists teachers in enhancing the quality of lectures, course content, and teaching methods. Consequently, this enhances learning performance, prevents disengagement, boredom, and reduces dropout rates.

The thesis aims to employ a camera for collecting behavioral data of learners. Subsequently, it utilizes various image processing techniques to extract and process frames from the recorded video. From these frames, the next step of the thesis involves the selection and identification of meaningful features to effectively assess the level of engagement. These features include eye open/closed status, facial expressions, body postures, etc. achieved through computer vision-based techniques.

As mentioned earlier, the features used to assess learner engagement include eye open/closed status, facial expressions, and body posture. However, the thesis opts for evaluating techniques based on eye open/closed status and facial expressions. The reason for this choice is that, despite numerous research works on these two techniques, there are still certain limitations when implementing them in real-world environments. Therefore, the thesis contributes to addressing some of these limitations.

The achieved results of the thesis are as follows:

- (1) Development of a learner engagement assessment framework based on eye open/closed status.
- (2) Introduction of an enhanced eye state detection technique. This technique is adaptable to various subjects with an automatically determined threshold, eliminating the need for a fixed threshold as seen in prior works.
- (3) Implementation of a technique for decomposing facial expressions into basic components. Given that learners' expressions in reality are

often a blend of fundamental emotions, and the training data is limited in terms of mixed emotions, decomposition is necessary to enhance accuracy for both recognition models and a detailed understanding of learners' emotions.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION.....	1
1.1 Problem statement	1
1.2 Necessity of the Thesis	2
1.3 Research Objectives	2
1.4 Object, Scope, Methodology	2
1.5 Contribution	3
1.6 Thesis structure	3
CHAPTER 2 OVERVIEW OF LEARNER CONCENTRATION ASSESSMENT AND FACIAL EXPRESSION RECOGNITION PROBLEM	4
2.1 Overview of learner concentration assessment.....	4
2.1.1 Significance of Concentration.....	4
2.1.2 Factors for Concentration Assessment.....	4
2.1.3 Concentration Assessment Methods	5
2.2 Facial Expression Recognition Problem.....	5
2.2.1 Facial Expression.....	5
2.2.2 Facial Express Recognition.....	6
2.2.3 Some approaches in assessing learner concentration based on facial expressions	6
CHAPTER 3 FACIAL EXPRESSION COMPONENT RECOGNITION.....	7
3.1 Evaluating concentration based on facial expression	7
3.1.1 Model architecture	7
3.1.2 Experimental evaluation	8
3.1.3 Concentration classification.....	8
3.2 Technique for Decomposing variations of a 3D model	9
3.2.1 Method	9
3.2.2 Deformation decay algorithm	10
3.2.3 Experimental results	10
CHAPTER 4 FACIAL EXPRESSION BEHAVIOR RECOGNITION	11
4.1 Evaluation of Concentration based on eye state	11
4.1.1 Theoretical Basis of the Evaluation	12
4.1.2 Evaluating Concentration based on Sleepiness.....	14
4.1.3 Concentration Assessment Algorithm	15
4.1.4 Experimental.....	17
4.2 Eye state detection technique	18
4.2.1 Theoretical foundation of the improved method.....	18
4.2.2 The enhanced technique for detecting eye open/closed states is based on variations in the eye aspect ratio.....	20
CHAPTER 5 CONCLUSIONS AND FUTURE WORKS.....	23

LIST OF SIGNS, ABBREVIATIONS

Acronym	Write full
AAM	Active Appearance Model
AEMS	Automatic Engagement Management System
AF	Activate Function
AU	Action Unit
CLAHE	Contrast-limited adaptive histogram equalization
CNN	Convolution Neural Network
DA	Data Augmentation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DT	Decision Tree
EAR	Eye Aspect Ratio
ECG	Electro Encephalo Graphy
EEG	Electro Encephalo Graphy
FACS	Facial Action Coding System
FER	Facial Expression Recognition
FERC-2013	Facial Expression Recognition Challenge
FFT	Fast Fourier Transform
HCI	Human Computer Interface
HE	Histogram Equalization
HMM	Hidden Markov Model
ITS	Intelligent Tutoring Systems
JAFFE	Japanese Female Facial Expression
KNN	K-Nearest Neighbors
LBP	Local Binary Pattern
LMS	Learning Management Systems
ReLU	Rectified Linear Unit
RF	Random Forest
SoTA	State-of-The-Art
SVM	Support Vector Machine

LIST OF FIGURES

Figure 2.1 Classifying Methods for Detecting Engagement [6].....	5
Figure 2.2 Some examples of AU [13].....	6
Figure 2.3 Architecture of a System for Assessing Learner Engagement Levels Based on EEG Signals [14].....	6
Figure 2.4 Architecture of a System for Assessing Engagement Levels Based on Computer Vision [17]	7
Figure 3.1 The architecture of the model for facial expression recognition	7
Figure 3.2 Algorithm Decomposition Flowchart.....	10
Figure 3.3 Some faces are randomly synthesized.....	11
Figure 3.4 The process of updating deviation error.....	11
Figure 4.1 Concentration Evaluation Flowchart.....	11
Figure 4.2 The time from eye opening to complete eye closure	12
Figure 4.3 Coordinates and frame rate of the eye.....	13
Figure 4.4 model estimates the level of sleepiness.....	14
Figure 4.5 Simulating the state of the learning process.....	14
Figure 4.6 The assessment learner concentration flowchart.....	15
Figure 4.7 Flowchart for calculating the study time of the learner	16
Figure 4.8 The result of estimating the level of sleepiness for a student	17
Figure 4.9 The result of evaluating the concentration of a student	17
Figure 4.10 Illustration of a student's concentration assessment results	18
Figure 4.11 Coordinates and frame rate of the eye.....	18
Figure 4.12 68 landmarks on face.....	19
Figure 4.13 Flowchart algorithm of Blink detection	21

LIST OF TABLES

Table 3.1 Results of the experimental run on the FERC-2013 dataset	8
Table 3.2 Emotion and its corresponding weight [21] [22]	8
Table 4.1 The correlation between sleepiness level and behavior [23]	12
Table 4.2 Rules for Determining Sleepiness Level	13
Table 4.3 Concentration Level	14
Table 4.4 Experimental results	21
Table 4.5 Results compared with other solutions	22

CHAPTER 1

INTRODUCTION

1.1 Problem statement

Globalization and technological advancements have been ongoing trends for many years, shaping a new global economy "powered by technology, fueled by information, and steered by knowledge". The integration of this new global economy has underscored the seriousness of the nature and role of education. Thus, learning has become immensely crucial, acting as a bridge to knowledge for each individual. The foundation of knowledge has revolutionized and transformed our society, completely altering how each person thinks, works, and lives. Furthermore, knowledge creates a harmonious living environment, bringing individuals closer together. Therefore, it can be asserted that learning is of paramount importance for each person; it is a prerequisite determining one's existence, integration, and development in society.

The learning process for each individual requires time and persistent effort to complete. In line with this perspective, scientists emphasize the significance of continuous effort in fulfilling learning tasks [1]. Various factors influence the learning outcomes of individuals, such as the educational environment, curriculum, class hours, infrastructure, and financial issues [2] [3] with a particularly critical factor being the learner's behavior [4] [5]. Thus, in any educational system, learner engagement and focus are the keys to success in learning.

To assess learner engagement traditionally, teachers conduct surveys to manually gather information through direct interviews with learners or by having learners self-report their levels of interest or boredom through predefined survey forms [6]. These methods are interventionist and human-dependent, suitable only for small class sizes due to their time and labor-intensive nature.

Today, technology has become an essential tool to support learners and educators for more effective educational experiences. Alongside this, Learning Management Systems (LMS) have emerged, playing a crucial role in university education teaching models [7]. This has opened up new research directions related to analyzing learner behavior within LMS to identify learning patterns for improving teaching methods for teachers and the learning process for learners.

On the other hand, the strong impact of the COVID-19 pandemic has affected various activities in people's lives, including education. This compelled education systems to shift from traditional to online learning formats

[8] [9] [10]. Although the pandemic has been controlled, and the situation has improved, many universities continue to adopt a blended learning approach, incorporating both traditional and online formats. This has led to an ever-expanding wealth of online teaching and learning data. A significant issue in online learning is the learner's interaction with their educational activities, given concerns about high dropout rates in online courses [6]. However, most current learning systems lack indicators/detectors of interaction levels [10]. Therefore, to understand the nature of the data, specifically teaching methods and learning behaviors, analyzing and assessing learner engagement has become one of the crucial indicators to measure the quality of teaching for teachers and the performance of learners.

Moreover, according to the author's viewpoint, an engagement assessment system can help education managers understand learner behavior comprehensively. This aids in improving the learning environment, infrastructure, training programs, and creates favorable conditions for learners to prevent issues such as dropping out. On the other hand, through learner behavior, it helps teachers enhance teaching methods to improve learner performance. Therefore, to contribute to enhancing the quality of teaching and learning, the author chose to research the topic "**Some Facial Expression Recognition Techniques for Assessing Learner Engagement**" for this thesis.

1.2 Necessity of the Thesis

Automated learning analysis is a highly essential issue in the field of education and training, especially with the increasing prevalence of numerous courses or online learning systems. This poses a challenge for teachers to monitor the learning behavior of students in large classes or in the same online class interface. Therefore, the research results of the thesis are a crucial foundation to help schools and teachers timely grasp the learning situation of students. From there, appropriate adjustments can be made to academic regulations, improve training programs, curriculum, lectures, and teaching methods to enhance the quality of education and learning outcomes for students, while quickly responding to new challenges in the education industry.

1.3 Research Objectives

The objective of the thesis is to investigate the problem of evaluating the concentration of learners. Subsequently, to address certain evaluation issues through image-based solutions using facial expression recognition techniques, inheriting existing results, and improving them to enhance the quality of the evaluation problem.

1.4 Object, Scope, Methodology

a) Research Object

The techniques related to the detection and extraction of facial features involve image and video processing.

Techniques for assessing engagement rely on analyzing the state of the eyes and facial expressions.

b) Research Scope

The thesis focuses on computer vision platform to develop techniques for evaluating concentration based on eye blink behavior and facial expressions of learners.

c) Research Methodology

The applied research methodology involves theoretical research, reviewing published works, and combining them with experiments. Issues related to algorithms and theories in image processing and computer graphics are addressed using computer software with inputs derived from real-world data. The process includes literature review, experimental setup implementation, result evaluation, and iterative improvements to enhance the quality.

1.5 Contribution

Proposing a technique to decompose facial expressions into basic components. During the learning process, learners constantly receive information from teachers, friends, environmental influences, etc. Therefore, the emotions of learners vary at each moment, being either mixed or individual, and the expressions may exhibit corresponding distortions.

Introducing a method to assess the level of learner concentration based on eye open/closed states using video data, and the analysis is conducted on each student throughout the learning process. This can assist teachers in making more informed adjustments to the curriculum, lectures, and teaching strategies to enhance learner performance.

Due to the unique characteristics of each person's eyes, the height and width of the eyes vary. Therefore, using a fixed threshold to determine the eye open/closed state is inappropriate and inaccurate. The thesis improves the eye state detection technique with an automatically determined threshold suitable for every individual to enhance detection accuracy.

1.6 Thesis structure

The structure of the thesis includes the following sections:

Chapter 1: This introductory chapter provides an overview of the significance of the research content. It introduces the problem statement and the approach to solving it. Additionally, it outlines the contributions of the thesis.

Chapter 2: This literature review chapter presents related issues regarding concentration assessment. It serves as a foundation for developing the proposed techniques in the subsequent chapters.

Chapter 3: Facial expression component recognition. This chapter details the technique for assessing learner concentration based on facial

expressions and proposes a technique to decompose facial expressions into basic components.

Chapter 4: Facial behavior expression recognition. This chapter discusses the technique for assessing learner concentration based on eye open/closed states. It also improves the eye state detection technique with an automatically determined threshold suitable for each individual.

Chapter 5: This chapter provides a summary of the research findings of the thesis and suggests future research directions.

CHAPTER 2

OVERVIEW OF LEARNER CONCENTRATION ASSESSMENT AND FACIAL EXPRESSION RECOGNITION PROBLEM

2.1 Overview of learner concentration assessment

2.1.1 Significance of Concentration

In education, concentration is often used to describe a learner's willingness to participate in the daily activities of the school, such as actively engaging in class, submitting required assignments, and following the teacher's instructions. According to [4] [11], there are three main factors contributing to the learner's concentration process, including behavior, perception, and emotion. Understanding the learner's behavior during class participation provides valuable information for the educational environment, such as:

- (i) **For educational managers:** Improving the learning environment, infrastructure, training programs, and creating favorable conditions for learners to prevent dropout.
- (ii) **For teachers:** Enhancing teaching methods, curriculum, and lectures to improve learner performance.

2.1.2 Factors for Concentration Assessment

Researchers have proposed solutions to monitor and assess the concentration of learners during their learning activities. In general, these solutions share similarities, and according to [4] [11], factors identifying engagement are categorized into three types: (1) Behavior, (2) perception, and (3) emotion.

2.1.3 Concentration Assessment Methods

Several studies on engagement detection have been published and classified into three main types: automatic, semi-automatic, and manual [6].

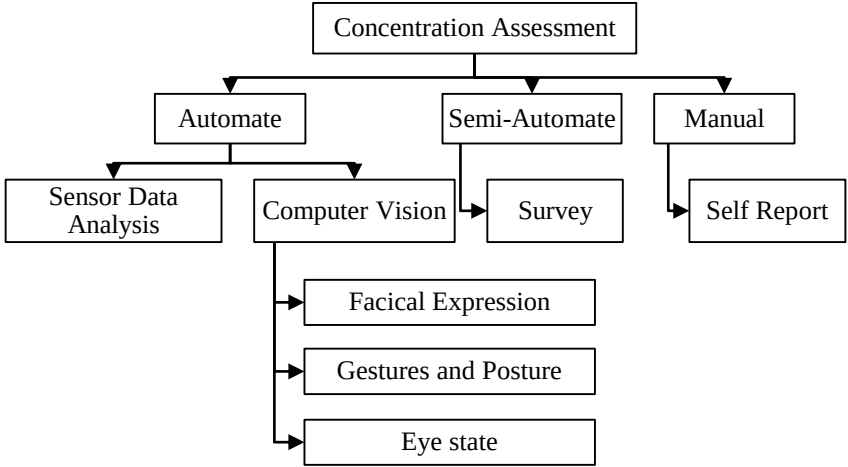


Figure 2.1 Classifying Methods for Detecting Engagement [6]

2.2 Facial Expression Recognition Problem

2.2.1 Facial Expression

2.2.1.1 Introduction

Ekman's [12] research reveals that there are six consistent facial expressions, corresponding to six basic emotions: joy, sadness, anger, surprise, disgust, and fear.

2.2.1.2 Action Unit

According to [12], there are sixty-four action units encoding facial movements when the face exhibits a specific emotion. In Appendix 1 of the thesis, a comprehensive list of the Facial Action Coding System (FACS) motion units is provided. Figure 2.2 illustrates some of the expression units on the face.



AU1
Inner Brow Raiser



AU2
Outer Brow Raiser



AU5
Upper Lid Raiser



AU9
Nose Wrinkler

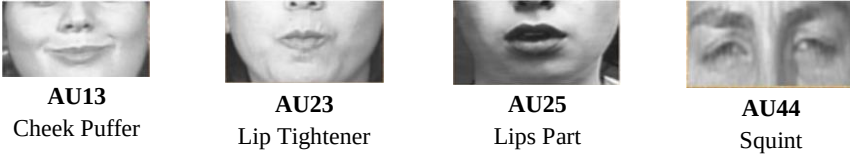


Figure 2.2 Some examples of AU [13]

2.2.2 Facial Express Recognition

The components on the face or facial landmarks are extracted to form a feature vector representing the geometric structure of the face. One of the most successful geometric-based methods is the Active Appearance Model (AAM). With appearance-based methods, image filters such as Gabor wavelets, Local Binary Patterns (LBP), are applied to the entire face or specific regions in the image to extract a feature vector.

2.2.3 Some approaches in assessing learner concentration based on facial expressions

In the field of education, analyzing learners' emotions has garnered significant attention over the past decades due to its connection with learning efficiency and educational outcomes. Assessing the engagement status of learners based on their emotions can be immensely beneficial for teachers. Therefore, the automated recognition of learners' emotional states during the learning process is highly anticipated. Facial analysis is one of the effective methods for automatically evaluating learners' emotional states in educational activities. Many scientists have proposed various methods for assessing learner concentration based on emotions, including two approaches: invasive and non-invasive.

Invasive methods: involve directly attaching devices to learners to measure physiological signals, such as electroencephalography (EEG) [14].

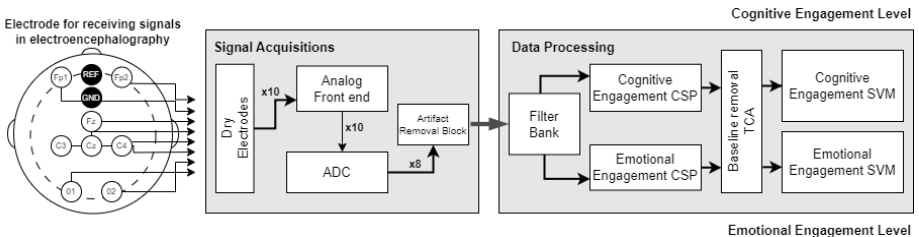


Figure 2.3 Architecture of a System for Assessing Learner Engagement Levels Based on EEG Signals [14]

Non-invasive methods: utilize computer vision techniques to analyze learning data. Some notable works have been proposed, such as [15] [16] [11]. The method involves a camera capturing images of learners in the classroom either in real-time or online. Subsequently, image and video processing techniques are employed to extract meaningful features, such as eyes and facial expressions, for analysis and evaluation through recognition algorithms or models.

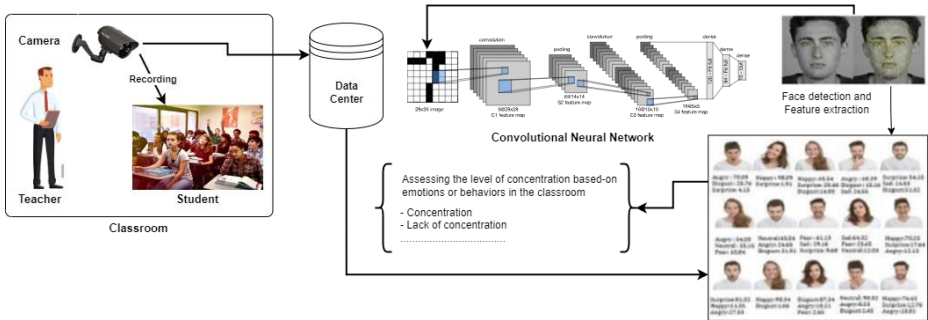


Figure 2.4 Architecture of a System for Assessing Engagement Levels Based on Computer Vision [17]

CHAPTER 3

FACIAL EXPRESSION COMPONENT RECOGNITION

3.1 Evaluating concentration based on facial expression

3.1.1 Model architecture

The architecture of the system is based on the VGGNet [18] and is described as follows

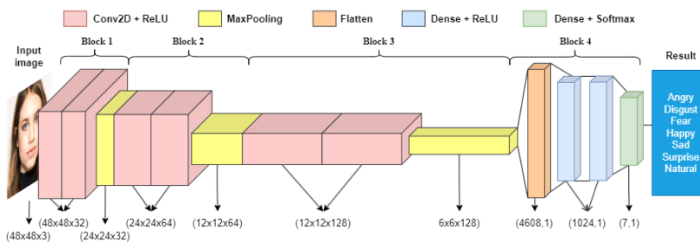


Figure 3.1 The architecture of the model for facial expression recognition

3.1.2 Experimental evaluation

3.1.2.1 Database

The FER-2013 dataset [19] was introduced at the International Conference on Machine Learning (ICML) in 2013 and is a standard dataset used in facial expression recognition challenges. This dataset comprises over 35,000 images, including 28,709 images for training and 7,178 images for testing, covering seven different facial expressions

3.1.2.2 Evaluation method and Experimental results

Table 3.1 Results of the experimental run on the FER-2013 dataset

	Precision	Recall	F1-score
Angry	0.56	0.58	0.57
Disgust	0.67	0.56	0.61
Fear	0.52	0.39	0.44
Happy	0.84	0.87	0.85
Sad	0.51	0.57	0.54
Surprise	0.78	0.78	0.78
Neutral	0.61	0.62	0.62

3.1.3 Concentration classification

a) Concentration Index (CI)

The engagement of learners can be measured by analyzing the emotional state of the learner (such as mood and emotion) based on facial expressions to estimate the overall level of engagement. The index measuring the concentration level is expressed as the following formula:

$$CI = DEP * EW \quad (3.1)$$

Where:

DEP (Dominant Emotions Probability): which is a value representing the predicted facial emotion from the recognition model.

EW (Emotion Weight): a value describing the level of a specific emotional state that reflects the concentration level of the learner at a particular moment. This value ranges from 0 to 1 and is represented by [20] [21].

Table 3.2 Emotion and its corresponding weight [20] [21]

No	Emotion	EW	CI
1	Neutral	0.9	(DEP * 0.9) * 100
2	Happy	0.6	(DEP * 0.6) * 100
3	Surprise	0.6	(DEP * 0.5) * 100
4	Sad	0.3	(DEP * 0.3) * 100
5	Disgust	0.2	(DEP * 0.3) * 100
6	Angry	0.25	(DEP * 0.25) * 100
7	Fear	0.3	(DEP * 0.3) * 100

b) Classification

According to works [20] [21], learner engagement is divided into three types: Not focused, low focus, and high focus. These are defined based on the emotional involvement of the learner during the lecture and are classified into three corresponding levels associated with the following basic emotions:

$$f(CI) = \begin{cases} \text{Not concentration, } CI < 20\% \\ \text{Low concentration, } CI \in [20\%, 50\%] \\ \text{High concentration, } CI \in (50\%, 100\%] \end{cases}$$

3.2 Technique for Decomposing variations of a 3D model

3.2.1 Method

3.2.1.1 3D model synthesis

Variations of the 3D model are assumed to be represented in the following form:

$$s = m + \sum_{i=0}^n \alpha_i * e_i \quad (3.3)$$

Where:

- s is the model variant to be aggregated;
- m is a model pattern, can be understood as a model of balance between distortions;
- e_i is the offset model corresponding to the deformation i ;
- α_i is the weight corresponding to deformation i ;
- n is the number of types of deformation of the model.

3.2.1.2 The problem Non-negative least squares

In this problem, the goal is to estimate a set of basic models, that is to compute m and e_i . We need the input data set of variants s and the weight values α_i corresponding to each variant. The input data set is modeled by the above formula as a matrix as follows

$$S = M + AE \quad (3.4)$$

Where:

- S is a set of input samples, S is a matrix $N \times L$.

- M is a matrix $N \times L$.
- A is a matrix $N \times n$ deformation weights.
- E is a matrix $n \times L$ are the compensation models.

If M is assumption, yes, the formula (3.5) will return:

$$S - M = AE \quad (3.5)$$

3.2.2 Deformation decay algorithm

Input : S, A

Output: m, E

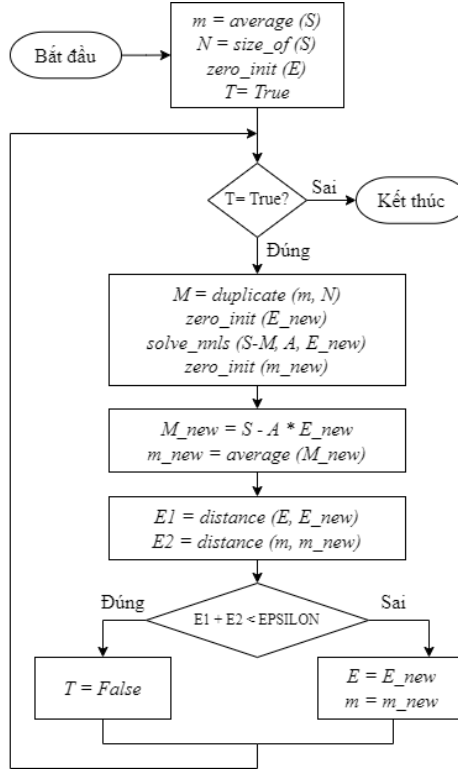


Figure 3.2 Algorithm Decomposition Flowchart

3.2.3 Experimental results

In the experimental program, the non-negative least squares method is taken from the Eigen library, which is specialized in calculating linear algebra.

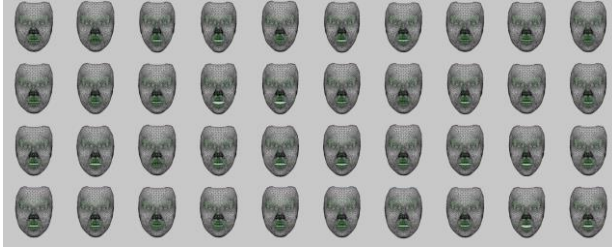


Figure 3.3 Some faces are randomly synthesized

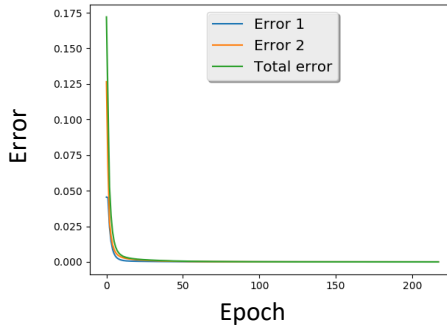


Figure 3.4 The process of updating deviation error

During the experiment, both the sample model and the complement model converge and are shown with deviation errors asymptotic to zero after more than 200 loops. The results are then applied to randomly synthesize the face model. At the synthesis step, each value in the expressive weight set is also taken at random in the segment $[0,1]$.

CHAPTER 4

FACIAL EXPRESSION BEHAVIOR RECOGNITION

4.1 Evaluation of Concentration based on eye state

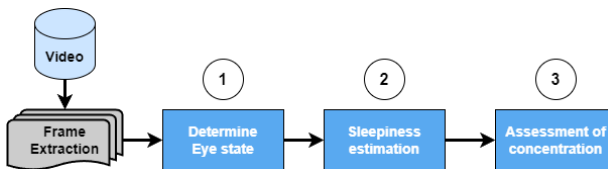


Figure 4.1 Concentration Evaluation Flowchart

4.1.1 Theoretical Basis of the Evaluation

The level of sleepiness is estimated based on various factors, such as frame rate and duration of eye closure. Each level corresponds to specific expressions, behaviors, and corresponding eye closure durations. Additionally, the speed of eye closure, whether fast or slow, is also considered. For an awake/alert person, eye closure and opening are rapid. For someone feeling drowsy, the eyes tend to close slowly [22].

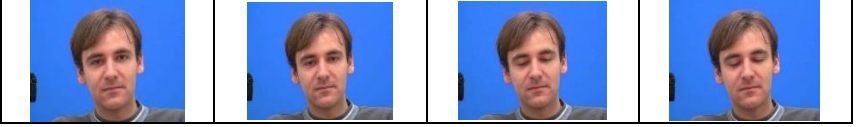


Figure 4.2 The time from eye opening to complete eye closure

In Figure 4.2 based on Talking Face1 data with a recording camera at 24fps, the complete eye closure state occurs at the fourth frame. Therefore, the time for complete eye closure is:

$$T_{eye_closed} = \frac{N_{frame} * 1000}{fps}$$

Where:

- N_{frame} the number of frames with the open eye to close eye
- fps (frames per second) the number of frames per second

Before estimating the level of sleepiness, a preprocessing step of the technique is to eliminate video segments with rapid eye closure ($T_{eye_closed} \leq 170ms$) see (4.1) and short duration of eye closure (duration < 1s), These segments, where the behavior occurs rapidly and does not affect the participation process, are considered as focused. The technique only focuses on video segments with slow eye closure ($T_{eye_closed} > 170ms$) and long duration of eye closure (duration $\geq 1s$) (see Table 4.1)

Table 4.1 The correlation between sleepiness level and behavior [22]

Status	Durations
Awake/Alert (Not sleepy)	<1s
Slightly sleepy	1-2s
Nodding off	2-3s or longer
Starting to doze off	$\geq 4s$

¹ https://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/talking_face/talking_face.html

Based on the degree and behavior in *Table 4.1*, the thesis establishes rules to determine the level of sleepiness as follows

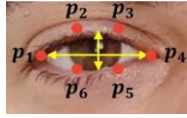
Table 4.2 Rules for Determining Sleepiness Level

Level	Condition
Awake/Alert (Not sleepy)	$EAR > 0.2$
Slightly sleepy	$EAR < 0.2$ AND $Duration \in [1s, 2s)$
Nodding off	$EAR < 0.2$ AND $Duration \in [2s, 4s)$
Starting to doze off	$EAR < 0.2$ AND $Duration \geq 4s$

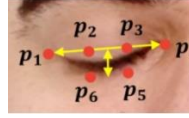
Where, EAR [23] The level of eye closure is calculated as the ratio of the height to the width of the eye as follows:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (4.2)$$

Where: $p_1, p_2, p_3, p_4, p_5, p_6$ is coordinates of the eye



(a)



(b)

(a) Frame rate of the eye when open; (b) Frame rate of the eye when close

Figure 4.3 Coordinates and frame rate of the eye

After calculating the value EAR , according [23], to determine whether the eye is closed or open, the author compares it to a threshold ($threshold = 0.2$) as follow:

$$f(EAR) = \begin{cases} \text{Eye Close,} & EAR < 0.2 \\ \text{Eye Open,} & \text{Otherwise} \end{cases} \quad (4.3)$$

Based on *Table 4.2*, the thesis defines 4 classes representing the level of sleepiness from 1 to 4, and two attributes to consider are EAR and $Duration$ (s). The model determines the level of sleepiness as follows:

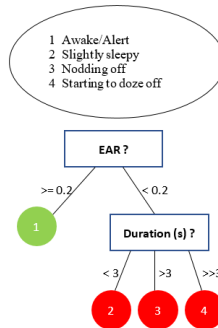


Figure 4.4 model estimates the level of sleepiness

4.1.2 Evaluating Concentration based on Sleepiness

Based on *Table 4.3*, the thesis constructs a concentration assessment scale for the system, consisting of 3 levels

Table 4.3 Concentration Level

STT	Level of sleepiness	Concentration Level
1	Awake/Alert (Not sleepy)	High concentration
2	Slightly sleepy	Low concentration
3	Nodding off	Not concentration
	Starting to doze off	

The evaluation method is detailed as follows:

Assuming the learning engagement process of the learner is simulated as shown in Figure 4.5, which includes various states occurring at different time points



Figure 4.5 Simulating the state of the learning process

Where

- T_{begin} : Start of learning time
- T_{end} : End of learning time



1 – Awake/Alert

2 – Slightly sleepy

3 – Nodding off



4 – Starting to doze off

Let t_1 is the total time for level 1, and t_2, t_3, t_4 is the respective total times for the remaining levels.

Let r_1, r_2, r_3, r_4 is the corresponding ratios for each level t_1, t_2, t_3, t_4

The ratios of the levels are determined:

$$\begin{aligned} r_1 &= \frac{t_1}{T_{end} - T_{begin}} & r_2 &= \frac{t_2}{T_{end} - T_{begin}} \\ r_3 &= \frac{t_3}{T_{end} - T_{begin}} & r_4 &= \frac{t_4}{T_{end} - T_{begin}} \end{aligned} \quad (4.4)$$

The concentration level is determined as follows:

$$x = \text{MAX}(r_1, r_2, r_3 + r_4) = \begin{cases} \text{High concentration, } x = r_1 \\ \text{Low concentration, } x = r_2 \\ \text{Not concentration, } x = r_3 + r_4 \end{cases} \quad (4.5)$$

4.1.3 Concentration Assessment Algorithm

a) **Algorithm 4.1:** Assessment of the learner's concentration

Input: Learning process videos

Output: the learner's concentration level

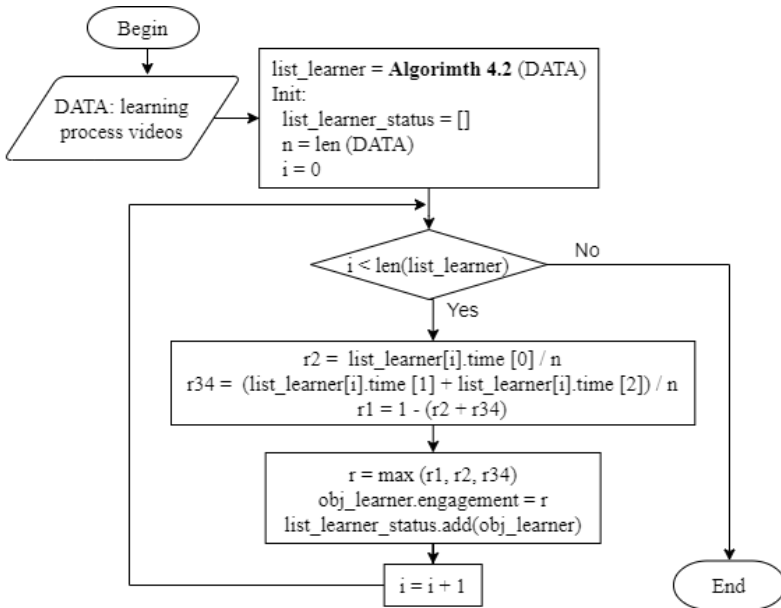


Figure 4.6 The assessment learner concentration flowchart

- b) **Algorithm 4.2:** Calculating the time of the learner's participation in the learning process

Input: Learning process videos

Output: The time of participation in learning of the learners at each level

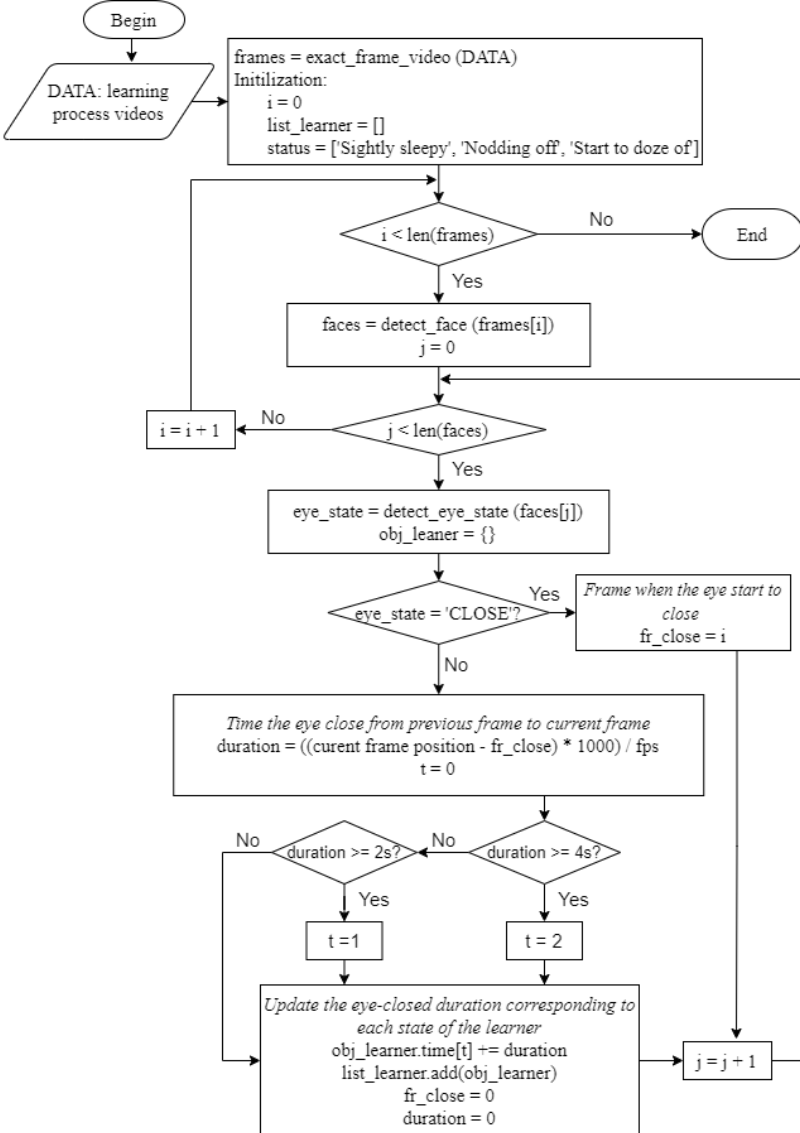


Figure 4.7 Flowchart for calculating the study time of the learner

4.1.4 Experimental

The thesis experiment tests the set objectives, including estimating the level of sleepiness based on eye opening/closing behavior and, finally, assessing concentration based on the level of sleepiness. Videos were recorded in two different classes during the break hour in the engineering department at Lac Hong University. The duration of each video experiment was approximately 10 - 15 minutes.

Case 1

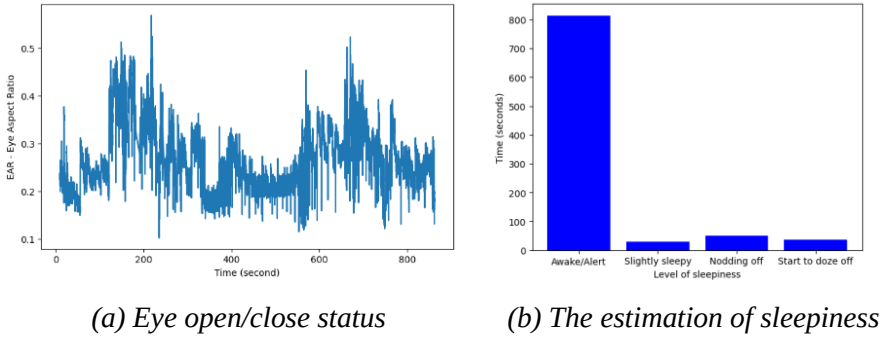


Figure 4.8 The result of estimating the level of sleepiness for a student

In which the system measures the detailed indices from *Figure 4.8* as follows:

- Awake/Alert: 813.99s take 87.65%
- Slightly sleepy: 29.22s take 3.15%
- Nodding off: 49.65s take 5.35%
- Start to doze off: 35.78s take 3.85%

The highest value among the levels is 87.65%, so based on *Table 4.3*, the system concludes that the subject is highly focused during the class hour as shown in *Figure 4.8*.

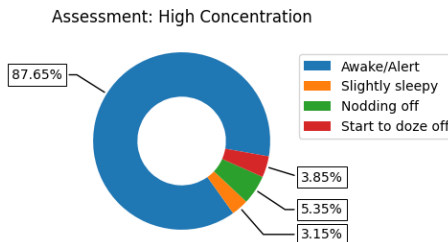
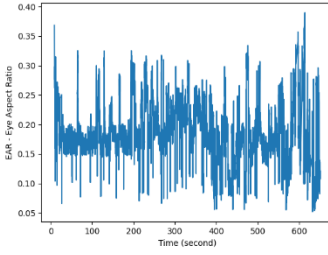
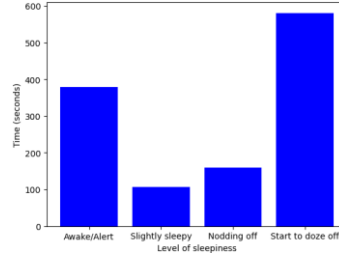


Figure 4.9 The result of evaluating the concentration of a student

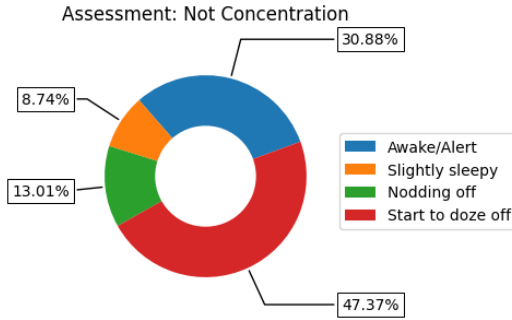
Case 2



(a) Eye open/close status



(b) The estimation of sleepiness



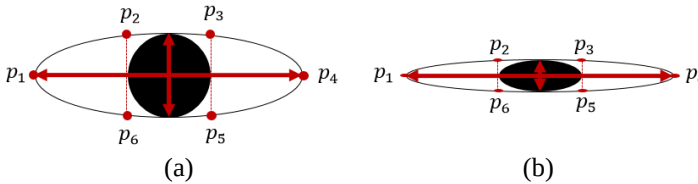
(c) Concentration Assessment Based on Sleepiness Level

Figure 4.10 Illustration of a student's concentration assessment results

4.2 Eye state detection technique

4.2.1 Theoretical foundation of the improved method

To calculate the eye frame rate, according to *Figure 4.11*, the person's eye is represented by a set of 6 points, starting from the left corner of the eye and marked clockwise. Each point represents a coordinate in 2D space with $p_1(x_1, y_1)$ and the rest correspond to points from p_2 to p_6 .



(a) Frame rate of open eye; (b) Frame rate of closed eye

Figure 4.11 Coordinates and frame rate of the eye

The width of the eye is the distance between points p_1 and p_4 determined by the Euclidean formula (4.6) according to Euclid.

$$D(p_1, p_4) = \sqrt{(p_1.x_1 - p_4.x_4)^2 + (p_1.y_1 - p_4.y_4)^2} \quad (4.6)$$

The height of the eye is the distance between pairs of points (p_2, p_6) and (p_3, p_5) determined by Euclidean distance similar to formula (4.6). The formula for the Eye Aspect Ratio (EAR) is determined as follows:

$$EAR = \frac{D(p_2, p_6) + D(p_3, p_5)}{2 * D(p_1, p_4)} \quad (4.7)$$

Where: $p_1, p_2, p_3, p_4, p_5, p_6$ To represent the corresponding coordinates on the eye

- With the left eye encompassing coordinates from p_{36} to p_{41} , and the right eye including coordinates from p_{42} to p_{47} in a clockwise direction (see *Figure 4.12*)

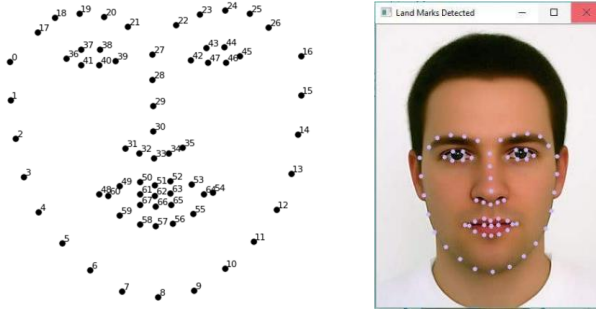


Figure 4.12 68 landmarks on face

The advantage of this method is its easy implementation, requiring no pre-training on eye open or closed data. The technique is suitable for real-time applications. However, it still has limitations as it needs to compare with a fixed threshold. This issue cannot generalize to all subjects because each person's eye features are different (some have large eyes, some have small eyes, etc.). Therefore, the thesis proposes an improvement to this method by automatically determining the threshold based on the eye features of each subject.

4.2.2 The enhanced technique for detecting eye open/closed states is based on variations in the eye aspect ratio

Due to the different sizes of eyes among individuals, for example, a person with larger eyes and another with smaller eyes will not have the same height and coordinate values. This leads to inaccuracies in determining the open/closed eye behavior when comparing the EAR value with a fixed threshold (where the eye is considered open or closed based on a predefined threshold). To improve this, the thesis proposes a solution to dynamically determine the open/closed eye behavior based on the variation in the frames containing the eyes for each individual, as follows

The frame containing the eyes in the normal state

$$EAR_{MAX} = \frac{D(p_2, p_6)_{max} + D(p_3, p_5)_{max}}{2 * D(p_1, p_4)} \quad (4.8)$$

The current frame containing the eyes

$$EAR_{CURRENT} = \frac{D(p_2, p_6)_{current} + D(p_3, p_5)_{current}}{2 * D(p_1, p_4)} \quad (4.9)$$

Assuming that the closing/opening eye behavior for a person occurs simultaneously in both eyes, the frame ratio containing the eye in the normal state and the current state is determined as follows

$$\begin{aligned} \overline{EAR}_{MAX} &= \frac{EAR_{MAX_L} + EAR_{MAX_R}}{2} \\ \overline{EAR}_{CURRENT} &= \frac{EAR_{CURRENT_L} + EAR_{CURRENT_R}}{2} \end{aligned} \quad (4.10)$$

Where: EAR_{MAX_L} is the normal frame rate of the left eye
 EAR_{MAX_R} is the normal frame rate of the right eye
 $EAR_{CURRENT_L}$ is the current frame rate of the left eye
 $EAR_{CURRENT_R}$ is the current frame rate of the right eye

Finally, to determine whether the eyes are closed or open, the technique compares the value $\overline{EAR}_{MAX}$ with $\overline{EAR}_{CURRENT}$ as follows

$$x = \overline{EAR}_{CURRENT} - \frac{\overline{EAR}_{MAX}}{2} = \begin{cases} \text{EYE CLOSE, } x < 0 \\ \text{EYE OPEN, } x \geq 0 \end{cases} \quad (4.11)$$

4.2.2.1 Flowchart algorithm of Blink detection+

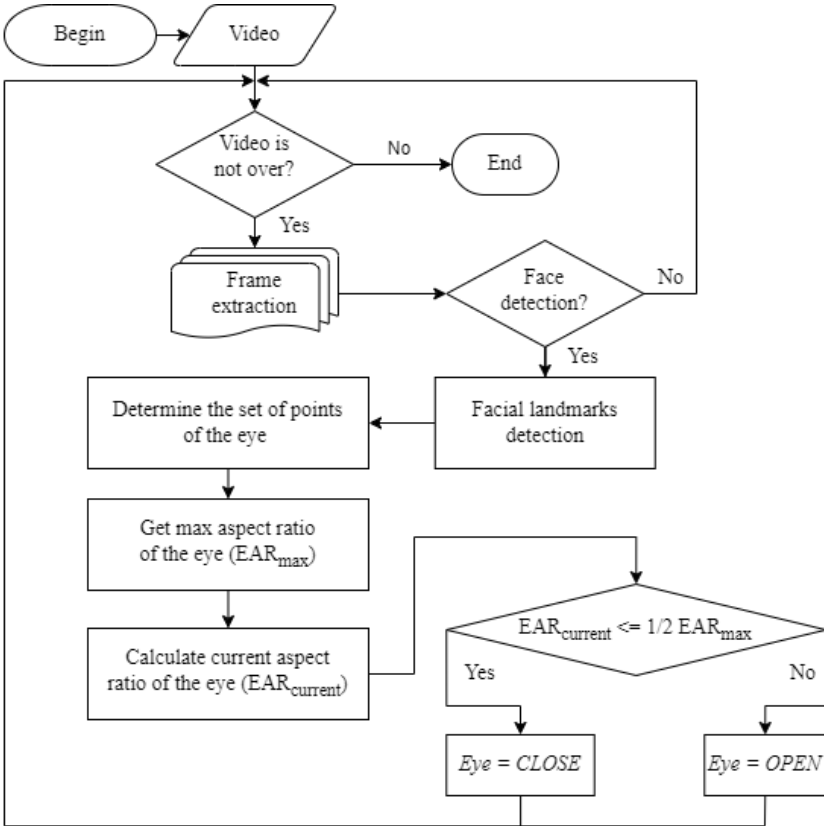


Figure 4.13 Flowchart algorithm of Blink detection

4.2.2.2 Experimental

Blink detection method compared to methods [23], [24] and [25] based on four datasets: **Talking Face**, **Eyeblink8**, **HUST-LEBW** và **ZJU**

Table 4.4 Experimental results

Dataset	SA	EC	TP	FP	FN	Pre	Recall	F1-Score
Talking face	1	61	58	3	4	0.9508	0.9355	0.9430
Eyeblink8	1	38	35	3	2	0.9211	0.9459	0.9330
	2	88	86	2	2	0.9773	0.9773	0.9770
	3	65	63	2	1	0.9692	0.9844	0.9770
	4	31	30	1	2	0.9677	0.9375	0.9520

Dataset	SA	EC	TP	FP	FN	Pre	Recall	F1-Score
	5	30	28	2	2	0.9333	0.9333	0.9330
	6	41	39	2	1	0.9512	0.9750	0.9630
	7	72	70	2	3	0.9722	0.9589	0.9660
	8	43	40	3	2	0.9302	0.9524	0.9410
						0.9528	0.9581	0.9550
HUST_LEBW	673	381	305	76	212	0.8005	0.5899	0.6790
ZJU	80	264	253	11	15	0.9583	0.9440	0.9510

Where: **SA**: Sample, **EC**: numbers of eye close, **Pre**: Precision

The chart illustrates that Blink detection+ performs effectively on datasets such as Talking face, Eyeblink8, ZJU, and less efficiently on the HUST_LEBW dataset. This dataset poses numerous challenges in a natural environment, such as wearing sunglasses, tilted images, etc.

Table 4.5 Results compared with other solutions

Methods	Dataset	Precision	Recall	F1-Score
[23]	Talking face	-	-	
Blink detection+	Talking face	0.9508	0.9355	0.9430
[23]	Eyeblink8	0.9430	0.9602	0.9520
Blink detection+	Eyeblink8	0.9528	0.9581	0.9550
[24]	HUST_LEBW	0.8295	0.4927	0.6180
Blink detection+	HUST_LEBW	0.8005	0.5899	0.6790
[25]	ZJU	0.9558	0.9367	0.9460
Blink detection+	ZJU	0.9583	0.9440	0.9510

Through the experimental implementation, the proposed method demonstrates higher accuracy compared to previously suggested methods. The method indicates that relying on landmarks is sufficient for accurately estimating eye open/close states. However, this approach still has limitations and performs less effectively in situations where the subject tilts their head significantly to the left or right. This leads to the method not fully recognizing the face, causing it to overlook some instances during detection.

CHAPTER 5

CONCLUSIONS AND FUTURE WORKS

The thesis has successfully achieved its research objectives aimed at constructing a system to support teachers in analyzing and evaluating the concentration of learners. The focus of the thesis is on the behavior and expressions of learners for assessment. The system functions as an assistant for decision-making in an automated manner. The thesis has completed the system development, data visualization, and has achieved the following research results:

- (1) Introduced a method to assess the level of learner engagement based on eye open/close behavior. The concentration assessment process is examined and analyzed throughout the learner's study. Initially, the thesis evaluates at individual time points and then synthesizes and calculates the ratio of each state concerning the entire learning process (as learners exhibit varying degrees of engagement at different times during learning).
- (2) Improved the eye open/close state detection technique with a dynamically determined threshold suitable for each individual. Due to the unique features of each person's eyes, the height and width of the eyes vary. Using a fixed threshold may not cover all cases.
- (3) Implemented a technique to decompose facial expressions into basic components to address the scarcity of training data. As learners constantly receive information from teachers, friends, environmental influences, etc., their emotions at each moment differ, and they may be a mixture or singular, resulting in corresponding distortions in expressions.

Future development orientations: While the thesis has proposed research results, there are still some inherent limitations. Monitoring learner behavior is subject to strict and rigorous requirements, and the evaluation needs to be examined and analyzed throughout the learning process.

- (1) The first limitation of the thesis includes the lack of monitoring other useful information, such as head orientation, hand gestures, yawning, etc. The technique has not covered all situations in learning.
- (2) The second is the inability to identify all facial landmarks for all learners due to partial obstruction or their heads facing a different

direction, not towards the teacher, leading to insufficient data for analysis in such cases and resulting in inaccuracies.

From these shortcomings, the author suggests the idea of combining multiple factors without relying on a fixed method, providing a basis for better decision-making, such as combining eye open/close to detect drowsiness, head orientation to detect sitting posture, and facial expressions for evaluation. Each individual method may be limited in certain cases when analyzing data.

Additionally, based on the initial results from the work [CT2], the thesis aims to explore the trajectory of head movement based on speech to assess the correlation between motion and content. This could provide a deeper insight into the relationship between emotions and learner behavior. Thus, the thesis also aims to realistically simulate how learners engage in the learning process by monitoring and reproducing movements.

Furthermore, this thesis has two areas that need improvement: (1) Since other authors have not tested their methods on standardized data but on separately collected data in their studies, comparing the effectiveness of the proposed methods in this thesis with other methods encounters difficulties; (2) the scale and experimental data are still "modest", so the effectiveness of the proposed methods needs further consideration.

REFERENCES

- [1] Michael ER Nicholls, Kellie M Loveless, Nicole A Thomas, Tobias Loetscher, and Owen Churches, "Some participants may be better than others: Sustained attention and motivation are higher early in semester," *The Quarterly Journal of Experimental Psychology*, vol. 68, no. 1, pp. 10-18, 2015.
- [2] Grasha, A.F., 2nd (Ed.), "Teaching with Style: A Practical Guide to Enhancing Learning by Understanding Teaching and Learning Styles," *Alliance Publishers: San Barnadino, CA, USA*, p. 167–174, 2002.
- [3] Richardson, M.; Abraham, C, "Modeling antecedents of university students' study behavior and grade point average," *J. Appl. Soc. Psychol*, p. 626–637, 2013.
- [4] Bui Ngoc Anh, Ngo Tung Son, Phan Truong Lam, Le Phuong Chi, Nguyen Huu Tuan, Nguyen Cong Dat, Nguyen Huu Trung, Muhammad Umar Aftab, Tran Van Dinh, "A Computer-Vision Based Application for Student Behavior Monitoring in Classroom," *Applied sciences*, vol. 9, no. 22, 2019.
- [5] Zheng, R., Jiang, F., & Shen, R., "Intelligent student behavior analysis system for real classrooms.," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing /ICASSP*), pp. 9244-9248. <https://doi.org/10.1109/ICASSP40776.2020.9053457>, 2020.
- [6] M. Ali Akber Dewan, Mahbub Murshed and Fuhua Lin, "Engagement detection in online learning: a review," *Springer Open*, vol. 6, no. 1, 2019.
- [7] Magdalena Cantabella, Raquel Martínez, España Belén, Ayuso Juan, Antonio Yáñez and Andrés Muñoz, "Analysis of student behavior in learning management systems through a Big Data framework," *Future Generation Computer Systems*, vol. 90, pp. 262-272, 2019.
- [8] Marko Horvat, Tomislav Jagušt, "Emerging opportunities for education in the time of COVID-19: Adaptive e- learning intelligent agent based on assessment of emotion and attention," *Proceedings of the Central European Conference on Information and Intelligent System*, 2020.
- [9] Jessie S. Barrot, Ian I. Llenares & Leo S. del Rosario , "Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines," *Education and Information Technologies*, 2021.

- [10] Shoroog Khenkar and Salma Kammoun Jarraya, "Engagement Detection Based on Analyzing Micro Body Gestures Using 3D CNN," *Computers, Materials & Continua*, vol. 70, pp. 2656-2677, 2022.
- [11] Pieter Vanneste, José Oramas, Thomas Verelst, Tinne Tuytelaars, Annelies Raes, Fien Depaepe and Wim Van den Noortgate, "Computer Vision and Human Behaviour, Emotion and Cognition Detection: A Use Case on Student Engagement," *MDPI Open Access*, 2021.
- [12] P. Ekman, W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Consulting Psychologists Press*, 1978.
- [13] Lucey P., Cohn J.F., Kanade T., Saragih J., Ambadar Z., Matthews I., "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA*, p. 94–101, 2010.
- [14] Andrea Apicella, Pasquale Arpaia, Mirco Frosolone, Giovanni Improta, Nicola Moccaldi, Andrea Pollastro, "EEG-based Measurement System for Student Engagement Detection in Learning 4.0," *Research Square*, 2021.
- [15] Sharma , P., Joshi, S., Gautam, S., Filipe, V., & Reis, M . J., "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning.," *CORR*, 190 9.12913. abs-1909-12913, p. [htt p://arxiv .org/abs/1909.12913](http://arxiv.org/abs/1909.12913), 2019.
- [16] Soloviev, V., "Machine learning approach for student engagement automatic recognition from facial expressions.," *Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics, Informatics and mechanics*, vol. 10, no. 2, pp. 79-86. <http://doi.org/10.5937/spsunp1802079s>, 2018.
- [17] Archana Sharma, Vibhakar Mansotra, "Deep Learning based Student Emotion Recognition from Facial Expressions in Classrooms," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6, pp. 4691-4699, 2019.
- [18] Chakradhar Pabba, Praveen Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert System Willey*, 2021.
- [19] Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski,W., Tang, Y., Thaler, D., Lee, D.H., et al.,

"Challenges in representation learning: A report on three machine learning contests. In: ICONIP.," *Springer*, p. 117–124, 2013.

- [20] P. Sharma, S. Joshi, S. Gautam, V. Filipe, and M. J. Reis, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," *arXiv preprint arXiv:1909.12913*, 2019.
- [21] Bouhlal Meriem, Habib Benlahmar, Mohamed Amine Naji, Elfilali Sanaa, Kaiss Wijidane, "Determine the Level of Concentration of Students in Real Time from their Facial Expressions," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 159-166, 2022.
- [22] Douglas M. Wiegand; Julie McClafferty; Shelby E. McDonald; Richard J. Hanowski, "Development and Evaluation of a Naturalistic Observer Rating of Drowsiness Protocol," 2009.
- [23] Soukupova, T. & Cech, J., "Real-Time Eye Blink Detection using Facial Landmarks," *21st Computer Vision Winter Workshop*, 2016.
- [24] Y. X. Z. C. L. M. Z. F. J. T. Z. a. J. Y. Guilei Hu, "Towards Real-Time Eyeblink Detection in the Wild: Dataset, Theory and Practices," *IEEE*, pp. 2194 - 2208, 2019.
- [25] Fogelton, A., and Benesova, W., "Eye Blink Detection Based on Motion Vectors Analysis," *Comput. Vision Image Understand*, 2016.