การรู้จำการแสดงออกทางสีหน้าสำหรับภาพภาษามือไทย

นางสาวเมย์ ทันดา เตย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558

FACIAL EXPRESSION RECOGNITION FOR THAI SIGN LANGUAGE IMAGE

Miss May Thandar Htay

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2015

Thesis Title	FACIAL EXPRESSION RECOGNITION FOR THAI SIGN LANGUAGE IMAGE
By	Miss May Thandar Htay
Field of Study	Electrical Engineering
Thesis Advisor	Assistant Professor Supavadee Aramvith, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.....Dean of the Faculty of Engineering
(Professor Bundhit Eua-arporn, Ph.D.)

THESIS COMMITTEE

.....Chairman
(Assistant Professor Widhyakorn Asdornwised, Ph.D.)
.....Thesis Advisor
(Assistant Professor Supavadee Aramvith, Ph.D.)
.....Examiner
(Assistant Professor Suree Pumrin, Ph.D.)
.....External Examiner
(Narut Soontranon, Ph.D.)

เมย์ ทันดา เตย์ : การรู้จำการแสดงออกทางสีหน้าสำหรับภาพภาษามือไทย (FACIAL EXPRESSION RECOGNITION FOR THAI SIGN LANGUAGE IMAGE) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. สุภาวดี อร่ามวิทย์, 47 หน้า.

งานวิทยานิพนธ์นี้แสดงขั้นตอนวิธีสำหรับการรู้จำการแสดงออกทางสีหน้า ที่ช่วยในการตีความหมายภาพภาษามือไทยโดยเพิ่มข้อมูลการแสดงอารมณ์เข้าไปด้วย ซึ่งการแสดงออกทางสีหน้านั้นสามารถแสดงให้เห็นถึงสภาพตอบสนองทางอารมณ์ภายในซึ่งบางสีหน้าที่เปลี่ยนไปอาจจะมีความคล้ายคลึงกัน โดยความคล้ายคลึงนั้นสามารถลดอัตราการรู้จำการแสดงออกทางสีหน้าได้ ดังนั้นขอบเขตของวิทยานิพนธ์นี้ คือ เสนอวิธีการรู้จำการแสดงออกทางสีหน้า ซึ่งประกอบไปด้วยสามขั้นตอน คือ การตรวจจับใบหน้า การแยกคุณลักษณะของใบหน้าและการรู้จำการแสดงออกทางสีหน้า บริเวณใบหน้าจะถูกตรวจจับโดยอัตโนมัติจากข้อมูลภาพขาเข้าโดยใช้ขั้นตอนวิธีไวโอลาร์-โจน์ ส่วนของใบหน้าจะถูกแบ่งออกเป็น ตาและปากโดยใช้วิธีการบรรทัดฐาน รวมกับคุณลักษณะจากรูปแบบเวกเตอร์ และขั้นตอนสุดท้าย คือ การรู้จำจะใช้ขั้นตอนวิธีการเพื่อนบ้านที่ใกล้ที่สุด และวิธีซัพพอร์ตเวกเตอร์แมชชีนมาใช้ในการจัดประเภทของการแสดงออกทางสีหน้า ในขั้นตอนวิธีของจะแบ่งชุดข้อมูลภาพออกเป็นสามกลุ่ม ได้แก่ ชุดข้อมูลภาพเชิงบวก ชุดข้อมูลภาพเชิงลบ และชุดข้อมูลภาพทั่วไป โดยจะเลือกภาพที่มีการเปลี่ยนแปลงการแสดงออกทางสีหน้าที่ไม่มีความคล้ายคลึงกันเป็นภาพขาเข้า โดยชุดข้อมูลภาพเชิงบวกภายใต้การแสดงออกเชิงบวก โดยจะเลือกภาพจากการขยับคิ้วขึ้นและการขยับมุมปากบนใบหน้า สำหรับชุดข้อมูลภาพเชิงลบ จะเลือกภาพจากการขยับคิ้วลงและการจีบปาก และชุดข้อมูลภาพทั่วไป จะเลือกภาพที่มีการแสดงออกทางใบหน้าที่เป็นปกติหรือเป็นธรรมชาติ โดยในงานวิทยานิพนธ์นี้ ได้เปรียบเทียบความถูกต้องของการรู้จำใบหน้าเดียว และการรู้จำใบหน้าที่รวมการแสดงออกทางสีหน้าที่เปลี่ยนไป ซึ่งผลการทดลองจะแสดงให้เห็นความถูกต้องของการจัดประเภทของปากจะมีค่าความถูกต้องที่มากกว่าส่วนอื่นๆ บนใบหน้า

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2558

5670554821 : MAJOR ELECTRICAL ENGINEERING
KEYWORDS: SIGN LANGUAGE, FACE DETECTION, FACIAL FEATURES EXTRACTION, FACIAL EXPRESSION RECOGNITION

MAY THANDAR HTAY: FACIAL EXPRESSION RECOGNITION FOR THAI SIGN LANGUAGE IMAGE. ADVISOR: ASST. PROF. SUPAVADEE ARAMVITH, Ph.D., 47 pp.

This thesis presents an algorithm for facial expression recognition to help the interpretation of the meaning of Thai sign language image by adding sentimental information. Facial expressions are the facial changes in response to the signer's internal emotional states. Some facial changes are similar. This similarity may cause the reduction in the facial expression recognition rate. We thus propose facial expression recognition framework in this thesis. The framework consists of the three components: face detection, facial features extraction and facial expression recognition. The face region is first automatically detected from input images by using Viola-Jones algorithm. The face components such as eyes and mouth are then extracted and normalized. Then, these features are combined to form feature vectors that will be used in the recognition. In the recognition step, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) are used to classify the facial expressions. In our algorithm, there are three datasets: positive, negative and neutral. For the dataset, we select the images that have the dissimilar facial changes as the input images. Positive dataset based on the positive expression that is identified by the facial changes: raising of the eyebrows and pulling of the mouth corner. For negative dataset, lowering of the eyebrows and puckering of the mouth are chosen to identify negative expression. For neutral dataset, expressionless facial changes are selected as the neutral expression. In this paper, we compare the classification accuracy by recognizing the single facial change and the combination of the facial changes for facial expression recognition. The experimental results show that the classification accuracy of the mouth is better than that of other parts of the face.

Department: Electrical Engineering Student's Signature .....
Field of Study: Electrical Engineering Advisor's Signature .....
Academic Year: 2015

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deep gratitude to everyone who has made it possible for me to successfully complete this thesis. First of all, I am deeply indebted to my advisor, Assistant Professor Supavadee Aramvith, Ph.D., for the great deal of effort she expended upon supervising me during my study in Chulalongkorn University.

My thanks also go to committee members, Assistant Professor Widhyakorn Asdornwised, Ph.D., Assistant Professor Suree Pumrin, Ph.D., and Narut Soontranon, Ph.D., for their contributing time, technical suggestions in the completion of this work.

I am grateful to thank to groupmates: Ksnokphan L., Adisorn P., Htoo Maung Maung, Watchara R., Sirawich S., Thipkesone B., and Chen S., for sharing their knowledge so many times. At this moment also I would like to say my gratitude to have Wai Phyoe Khaing who always supports me in every step.

A word of thanks is not enough for members in my family. They gave me motivation and patience while for me to complete this thesis. I would like to dedicate this work and express my heartfelt appreciation to them.

I would like also to send my thanks to Ministry of Science and Technology in Myanmar, for directing me on my study plan at Chulalongkorn University.

Finally, I would like to send my gratefulness to the staffs of ISE and AUN/SEED-Net, for offering me the chance of pursue my Master degree. Without their financial assistance, care, and help, my study at Chulalongkorn University could not be processed.

This research has been supported in part by the Collaborative Research Project entitled Wireless Video Transmission, JICA Project for AUN/SEED-Net, Japan.

CONTENTS

Page

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Motivation and Significance of the Research Problem

The deaf people use sign language as a communication media in hearing-impaired or deaf community. Sign language conveys information through three main parts: hand gestures that employ hand shape and motion to express a word or meaning, facial expressions which modify the meaning of a sign, and finger spelling that spells out the words gestural in the local verbal language. The deaf people can only communicate by using visual. Therefore, a deaf person still has difficulties not only in accessing information and knowledge but also in communicating with a normal person who does not understand sign language.

To overcome these difficulties and to be able to communicate with the normal people, the deaf people need an interpreter. However, an interpreter is not the answer because of the cost and the limited availability of trained personnel. Therefore, Sign Language Recognition (SLR) system plays an important role to interpret sign language in hearing-impaired or deaf community. Thai Sign Language (TSL) is one of the sign languages that are uniquely used by deaf Thais. In Thailand, to narrow down the communication gap between the deaf people and the normal people, many researchers focus on Sign Language Recognition (SLR). SLR can support the deaf people by enabling the development of systems for Human-Computer Interaction (HCI) in sign language and translation between sign and spoken language.

In Human-Computer Interaction (HCI) applications, facial expressions directly link to human emotions such as happy, sad, angry, fear, surprise and so on [7]. The facial expressions we use while doing a sign will affect the meaning of that sign. For example, if we sign the word “fear,” and add an exaggerated or intense facial expression, we are telling our audience to be "very fear." This principle also works when making "surprise" into "very surprise," or "happy" into "very happy". If hand gestures and finger spelling change slightly, it may become a different sign. Facial expressions accompany the meanings of signs by using human emotions. Therefore, to

understand better sign languages, facial expression recognition systems are quite important in Sign Language Recognition (SLR).

Early research basically revolves around understanding of hand gestures recognition and finger spelling recognition in SLR. Saengsri et al. [2] proposed Thai sign language recognition system by using finger-spelling recognition. In their research, they used a data glove to acquire specifics of all five fingers. And, they provided hand movement by using a motion tracker device. From the data glove and motion tracker, they obtained the advantages of the accurate data. Suksil and Chalidabhongse [3] segmented face and hands by skin color model in YCbCr color space and labelled face and hands' initial positions for further tracking by Haar-like feature. In tracking, they used object hypothesis and template matching to solve face and hands occulation, usually occur in sign language. Their results were good in hand gesture recognition. To enhance the results, their system allowed the flexibility of incorporating additional techniques.

Soontranon et al. [22] presented sign language recognition system by localizing and tracking face and hands. To find the most suitable one without parametric model approach, they calculate preliminary evaluation over various color spaces. They focused on to better model the skin color and to lower the complexity of the detection algorithm by using the elliptical model on CbCr. In their system, they segmented the skin color regions from the sign language input videos and then detected the interested hands and facial features by using skeleton features and luminance differences respectively. After detecting process, the next step was tracking. To match its own blob, each blob determines the search region and calculates MMSE (Minimum Mean Square Error) in the tracking step. They used the block matching method between current and previous frame. The experiments show that their system was able to detect and track hands and face from the sequences of the sign language videos. However, these researches mainly focused on finger-spelling and hands movements. Facial expression did not take into account in their systems. Therefore, research needs to turn around understanding of non-manual features recognition in SLR to modify the meaning of manual features.

In this research, we will propose an algorithm that recognizes the facial expressions in Thai sign language. We believe that our research can improve the

performance of facial expression recognition systems that will be helpful in deaf community of Thailand.

1.2 Research Contributions

The main contribution in this research is to propose facial expressions recognition that can help to interpret the meaning of TSL. To localize and extract the face regions of the images, face detection is a first-step in the facial expressions recognition. For face detection, Viola-Jones algorithm automatically detects the face region for the input images of the TSL videos. After face detection process, the detected face, eyes including eyebrows and mouth are extracted.

In the facial features extraction, all of the extracted face regions are normalized to remove the variation effects in the scale of the images. After obtaining the normalized features, these features combine to form a feature vector.

For the facial expression recognition, KNN algorithm is trained to predict the facial expressions according to feature vector of the facial features extraction. KNN is a simple algorithm for facial expression recognition. In training, the images are averaged and normalized in a given class to generate a template for this class. After training, KNN algorithm matches the input image with the closest template by using Euclidean distance. Euclidean distance measures the dissimilarity between the input image and the closest template.

1.3 Objectives

1. Analyse the facial expressions of a facial image by combining and extending the existing algorithms.
2. Develop an algorithm for the facial expression recognition for Thai sign language video.

1.4 Scope of the thesis

1. Propose a framework that can help the interpretation of the meaning of Thai sign language by adding emotional information from facial expressions.

1.5 Research Procedures

1. Review literatures related to Thai sign language, face detection, facial features extraction, facial expression recognition, and existing sign language recognition systems.

2. Select the input images from Thai sign language videos and group these input images into different categories.
3. Select the most suitable method for face detection from existing face detection methods.
4. Implement facial features extraction method.
5. Create training dataset and train for facial expression recognition by using that dataset.
6. Implement facial expression recognition which will use trained classifier as an input.
7. Verify the performance of the proposed facial expression recognition system.
8. Summarize the results, analysis, and conclude the performance of the proposed system.
9. Take proposal examination.
10. Submit the paper.
11. Write the thesis.
12. Take thesis defense examination.

1.6 Thesis Organization

This thesis is organized into five chapters including this chapter. The following paragraphs provide brief descriptions of the remaining chapter of this thesis.

Chapter 2 discusses some basic of face detection, facial features extraction and facial expressions recognition. Some characteristics of TSL are also mentioned in this chapter.

Chapter 3 presents the proposed algorithm. Face detection is the first stage using Viola-Jones algorithm to detect the face. The next stage is to extract the facial features by using geometric-based method. For the recognition in the last stage, KNN algorithm classifies the facial expressions.

Chapter 4 presents the experiment and results of proposed method.

Chapter 5 includes conclusions and future works of the research.

CHAPTER II

BACKGROUND

2.1 Thai Sign Language

Sign language plays an important role in hearing-impaired or deaf community because they use sign language as a communication media not only for expressing thoughts but also for exchanging information. There are different sign languages all over the world, just as there are different spoken languages. English-spoken countries are Australia, United Kingdom and United States of America. Their own sign languages are Australia Sign Language (Australia) [11], British Sign Language (BSL) [9], and American Sign Language (ASL) [8]. Their own sign language represents the same English word by different signs. Similarly, Thailand also has its own sign language. Thai Sign Language (TSL) is the sign language that uses the combinations of hand gestures, finger spelling and facial expression to communicate without using sound in deaf Thais.

In TSL system, signs can be differentiated into five categories: Pantomimic signs, Imitative signs, Metonymic signs, Indicative signs, and Initialized signs [1]. Pantomimic signs convey meanings of the signs by using hand gestures simultaneously. Imitative signs convey particular meanings instead of the overall meanings. For example, these signs convey the meaning of both "car" and "car driving" by signing of hand holding steering wheel of the car. Metonymic signs convey meanings of the signs by presenting the signs as a reference result. For example, the saluting sign refers to "soldier." Indicative signs convey meanings of the signs by pointing the finger to the reference object. For example, the signers just simply point their index finger toward their chest to represent themselves. In a spoken/written language, Initialized signs convey meanings of the signs by initializing hand-shape corresponding to the first letter of the word.

Since 1950s, Thailand provided educational supports to deaf Thais. However, the place of deaf Thais still has limitations in the society because deaf Thais lack the tools such as dictionary that helps them in their communication. In Thailand, American-

trained Thai educators introduced deaf schools since 1950s. Therefore, some parts of Thai Sign Language (TSL) related to American Sign Language (ASL). Woodward [10] analysed potential cognates between American Sign Language (ASL) and Modern Standard Thai Sign Language (MSTSL). There was a 52% rate (47/90 pairs) of possible cognates between ASL and MSTSL. All of these percentages indicated that ASL and MSTSL should be classified as distinct languages that are closely related historically and that belong to the same language. Thus, there was no systematic approach in TSL because ASL influenced Thai signs in school for deaf Thais.

Therefore, the deaf teachers were well-educated to be more systematic the study of signing in TSL. The dictionary plays an important teaching tool in the development of TSL. In 1979, a dictionary name was "Signing Exact Thai & English" that released by Sethsatian School. It consists of Thai words (3,000), English words (1,860), illustrations (2,000), and notation together with explanations in Thai and English. By using the notation system with hand-shapes, Charles Reilly and Manfa Suwanarat proposed the project for compination of the TSL in 1980. Their project result was "The Thai Sign Language: A Model Dictionary by Handshapes". In 1986, this dictionary became a model of the "Thai Sign Language Dictionary". To support the educations for the deaf Thais, National Association of the Deaf in Thailand (NADT) becomes as an active organization for deaf Thais in 1983. NADT revised and extended TSL dictionary in 1990. Instead of replacing TSL with sign languages from developed countries, this directed research efforts into their own sign language for deaf people in Thailand.

Supavadee Aramvith and Teeranoot Chauksuvanit proposed a development of an online electronics TSL (eTSL) dictionary from analysis of the history and features of TSL dictionary. This eTSL dictionary included finger spelling system as well as hand gestures. Therefore, this dictionary supports more TSL vocabularies to communicate for deaf Thais. This dictionary system will employ Sign Language Recognition (SLR) to narrow down the communication gap between the deaf people and hearing people. Sign Language Recognition (SLR) system interprets the signs by conveying the information through three main parts: hand gestures recognition, finger spelling recognition and facial expressions recognition.

Electronics Thai Sign Language Communication System is the first project in Thailand to address the accessible communication of Deaf to the society and to facilitate

communications among Deaf and between Deaf and hearing people supported by NBTC (2011-2012). The prototype composes of video and virtual video communication, Thai sign language recognition, and web based dictionary. It is expected that the deployment of the system in public domain and to the Deaf community would help realize the goal of accessible communication for all. Currently we are seeking funding support to develop ETSL based mobile applications. H.264, HEVC and propose enhancements to make better quality video communications possible.

2.2 Face Detection

In the facial expression recognition systems, face detection is a first-step to localize and extract the face regions of the images. It also has several applications in areas such as crowd surveillance, facial expression recognition systems, face recognition systems, and computer vision communication. In recent years, face detection is a very important quick developing research area. The face detection in real time system still has problems because of varying pose, presence or absence of structural components, facial expression, occlusion, image orientation, lighting condition, and low resolution images.

2.2.1 Neural Network Based Face Detection

Rowley et al. [14] proposed a neural network algorithm for the upright face detection in the grayscale images. Their study focused on the improvement of the performance over a single network. They use a bootstrap algorithm to select the negative examples. This bootstrap algorithm is easy to collect nonface training examples for training. There are two steps in their algorithm. First is to apply neural network filters to an image and second is to use an arbitrator for combining the outputs. The neural network filters examine each face locations of the images at several scales. From individual filters, the arbitrary combines the detections and excludes the overlapping detections. Figure 1 shows the basic algorithm used for face detection of their proposed algorithm.

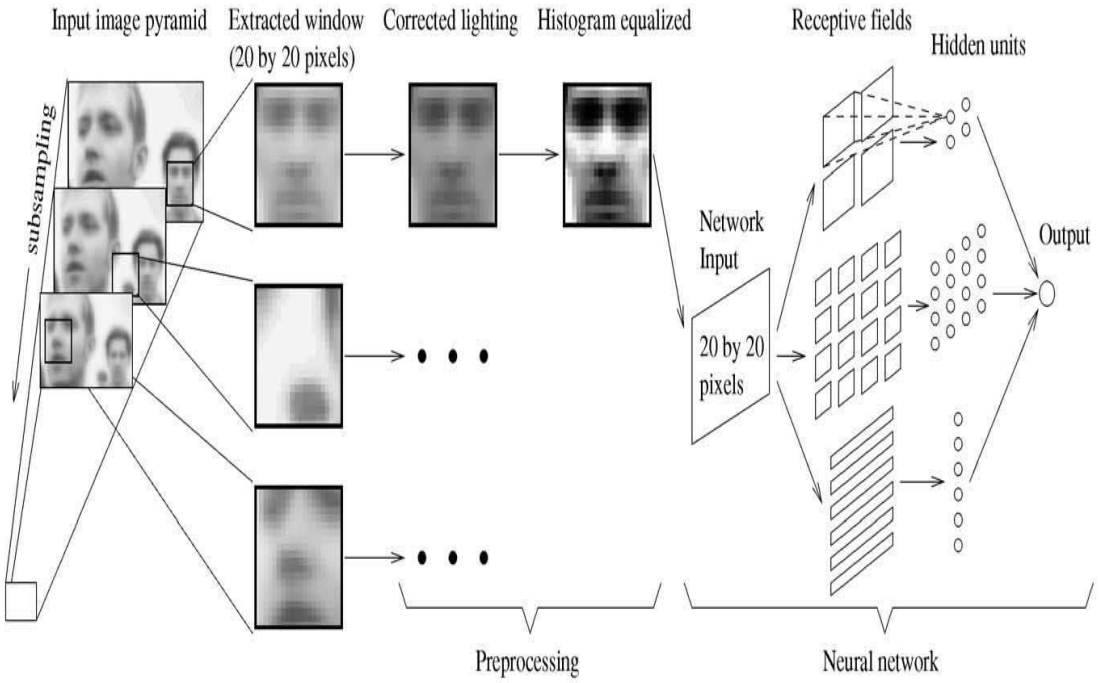


Figure 1. The basic algorithm used for face detection [14]

As an input image, a filter receives an image that has a 20x20 pixel region and generates an output by signifying whether face is present or absent ranging between 1 and -1. If the image is larger than the window size, the size of the input image is repeatedly reduced by subsampling. To detect the presence of a face in the window, the window is passed over a neural network.

To make equal the intensity values, the preprocessing step is applied by using the linear function. By the linear function, the overall brightness of each part of the window will be approximate. To compensate for a variety of lighting conditions, this linear function subtracts the overall brightness from the window. After compensating the lighting conditions, histogram equalization is computed for the pixels in an oval region of the window.

To detect the local features for the face detection, there are three types of hidden units in the neural network. The square receptive fields allow the hidden units for detection these local features as left eye, right eye, nose, or mouth corners. The hidden units with horizontal stripes detect the local features such as mouths or eyes pairs. Then, the arbitrary combines the detections and excludes the overlapping detections from individual filters.

Their algorithm can detect the faces with the comparable performance and false positive rates. On a 200 MHz R4400 SIG Indigo 2, their algorithm can process a 320x240 pixel image within 2 to 4 seconds.

2.2.2 Component Based Face Detection

Heisele, Bernd, et al. [15] presented a component based algorithm for the frontal and near frontal face detection in the grayscale images. By using 3D head models, their study focused on automatically learning components. In their system, there are two levels of Support Vector Machin (SVM) classifiers. First level is independently to detect the components of the face with component classifiers. In the second level, a single classifier matches the geometrical configuration of the detected components with the geometrical model of the face. Figure 2 shows the system overview of the component based classifier using four components.

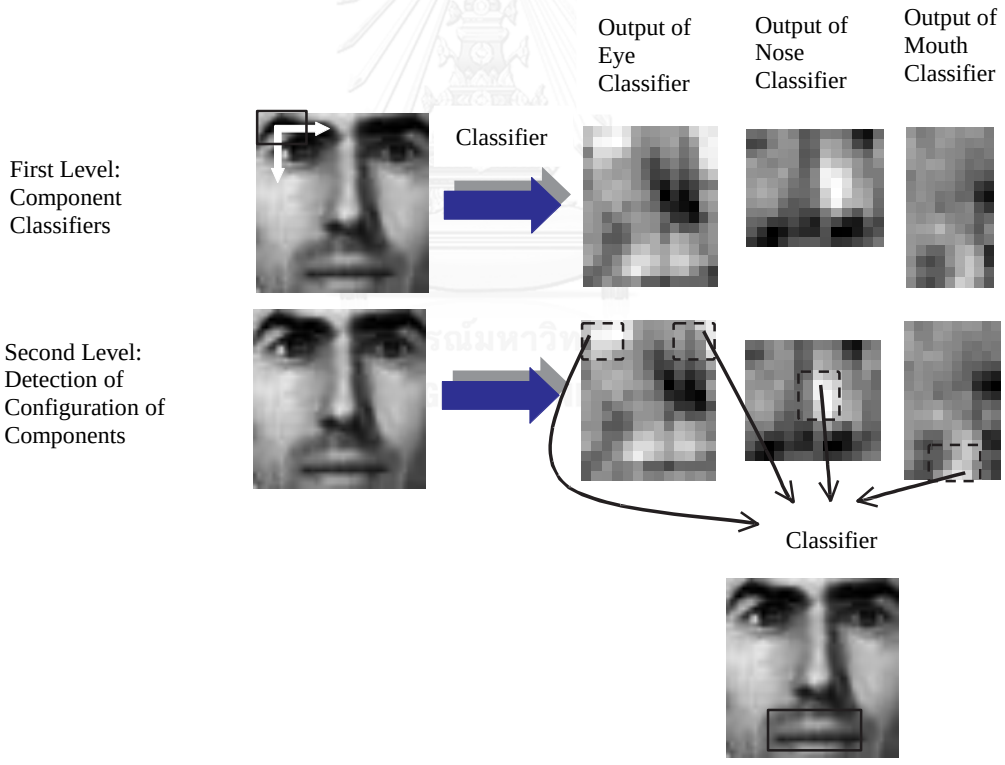


Figure 2. System overview of component based classifier using four components [15]

With component classifiers, the components of the face are independently detected in the first level. Each of SVM classifier was trained on a set of non-face patterns and a set of facial components that were automatically extracted from synthetic

58x58 face images. By linearly combining the results generated from the component classifiers, the final face detection was performed by the geometrical configuration classifier. In their system, the classification performance of the component based system is significantly better compared to the classification performance of a whole face detection system.

2.2.3 Real Time Face Detection

To detect the face in real time face detection, Viola and Jones [4]-[5] introduced the fast and efficient face detection method. In Viola-Jones face detection [4]-[5], all Haar-like features consist of three basic types of rectangle features in figure 3: two-rectangle, three-rectangle and four-rectangle features. The value of a two-rectangle feature is the difference of the summed pixels between black and white rectangles. A three-rectangle feature calculates the difference between the sum within two outside white rectangles and the sum in a centre black rectangle. Finally a four-rectangle feature computes the difference between the sums of two rectangle pairs in the diagonal.

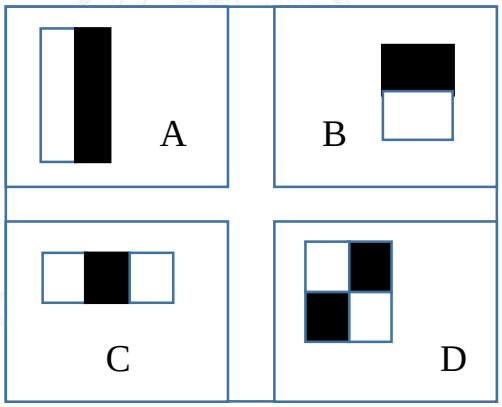


Figure 3. Basic Haar-like feature types [5]

A feature is evaluated by the difference between the summed pixel values of black rectangle and that of white rectangle.

Rectangle features:

$$f(x) = \sum(\text{pixels in black area}) - \sum(\text{pixels in white area}) \qquad (1)$$

The most important characteristic is that the integral image computes rapidly the Haar-like features at all scales in constant time.

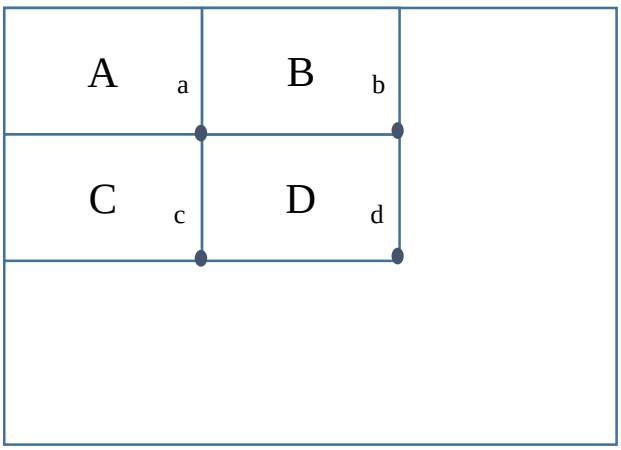


Figure 4. Integral image [5]

In figure 4, the integral image calculates the sum of all pixels by using only four values at the corners of the rectangle. Integral image computes the sum of the pixels within D with four array references. The value of the integral image at location a is the sum of the pixels in rectangle A. The value at location b is A+B. The value of the integral image at location c is A+C and the value at location d is A+B+C+D. The sum within D can be computed as d + a – b – c = (A+B+C+D) + A – (A+B) – (A+C) = D.

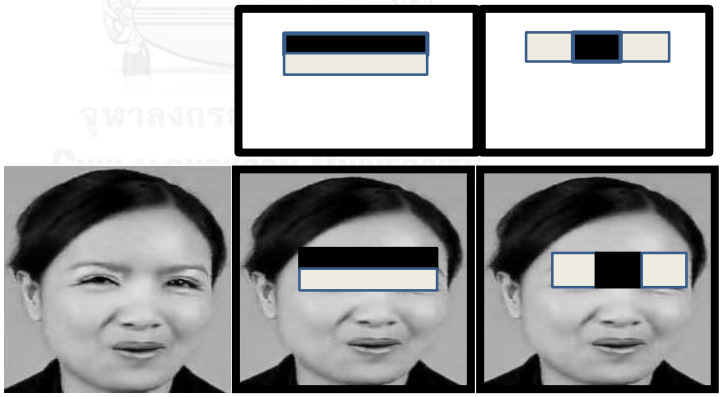


Figure 5. Features selected by Adaboost [5]

The features locate at any area within a scanning image window. A 24x24 window results over 160000 features. For fast recognition, the machine learning process must exclude the large number of rectangle features and focuses on a small set of best features among the large number of rectangle features. Adaboost [6] is a machine learning algorithm which selects only the best features among all these rectangle features and trains the classifier. In figure 5, the top row shows the two good features

selected by Adaboost. Bottom row overlays these two features on a typical training face. The first selected feature focuses on the property that the region of the eyes is darker than the region of the nose and cheeks. The second selected feature relies on the property that the eyes are darker than the bridge of the nose. These rectangle features (also known as weak classifiers) cannot classify the image. However, these rectangle features form a strong classifier by combining with other rectangle features.

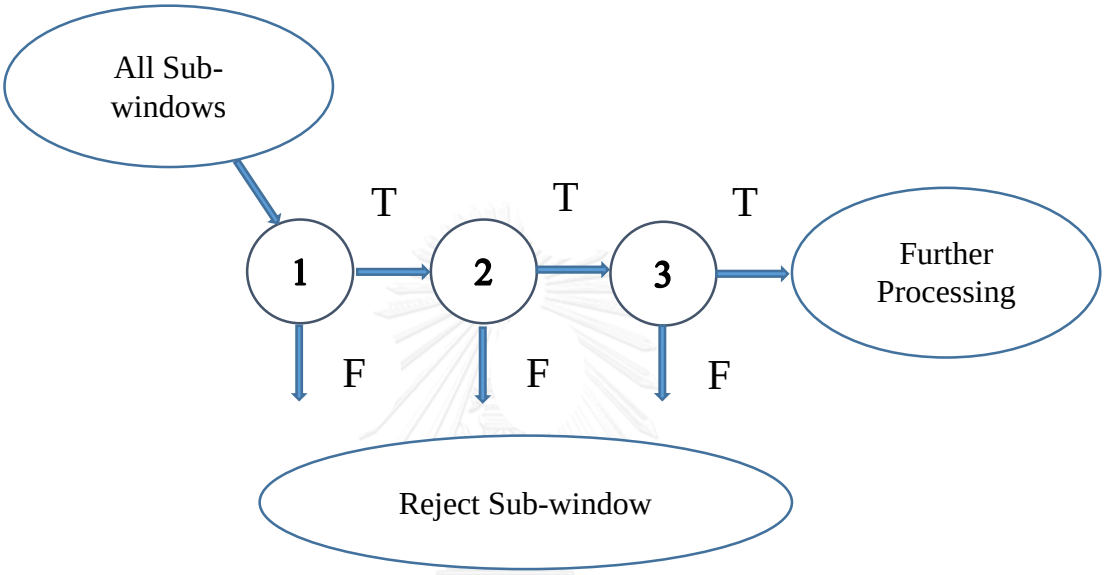


Figure 6. Simple diagram of Adaboost cascading [5]

In figure 6, the cascade classifier is made up of stages. Each stage is composed of an AdaBoost strong classifier. Each stage is used to determine if a sub-window contains a face or non-face. If a sub-window fails in the first stage, it is rejected as non-face. If a sub-window passes in the first stage, it is applied to the second stage as face and the process is continued by adding stages until the target detection is met.

2.3 Facial Features Extraction

Facial features extraction extracts the facial features such as eyebrows, eyes, and mouth as well as their spatial relationships from the detected face region. It generates the facial features to be used in facial expression recognition. Therefore, it plays an important role in facial expression recognition because the facial features can make a better recognition of the facial expression. The result of the facial features extraction is a set of features called a feature vector. This feature vector classifies the facial expressions such as question, negation, neutral, happy, sad, angry, fear, surprise

and so on. There are three types of feature extraction methods: color based, texture based and geometric based methods.

2.3.1 Geometric Based Method

To extract the facial features, the geometric-based method uses sizes and relative positions of the important components from the detected face region. This method detects the edges and direction of important components and then builds feature vectors from these edges and directions. In Adaboost [6] method, Haar-like features [5] change the grayscales distribution into the feature. Haar-like features represent the features of the face by the difference between the summed pixel values of black rectangle and that of white rectangle in figure 7. Adaboost [6] extracts only the best features among all these rectangle features and then trains these best features to form a feature vector. Local Binary Pattern (LBP) [12] method divides every detected face region into blocks. Each block corresponds to central pixel. LBP method compares grayscale values of the central pixel with every pixel of all eight neighbours in the block. If the grayscale value of the central pixel is greater than the grayscale value of the neighbour pixel, the value of the neighbour pixel is zero. If the grayscale value of the central pixel is less than the grayscale value of the neighbour pixel, the value of the neighbour pixel is one. In figure 8, the result of the LBP method is binary 00011110. Therefore, a binary string represents every pixel in the block. This method builds a histogram for each block. To form a feature vector, this method combines the histograms. This feature vector classify the facial expressions such as question, negation, neutral, happy, sad, angry, fear, surprise and so on.

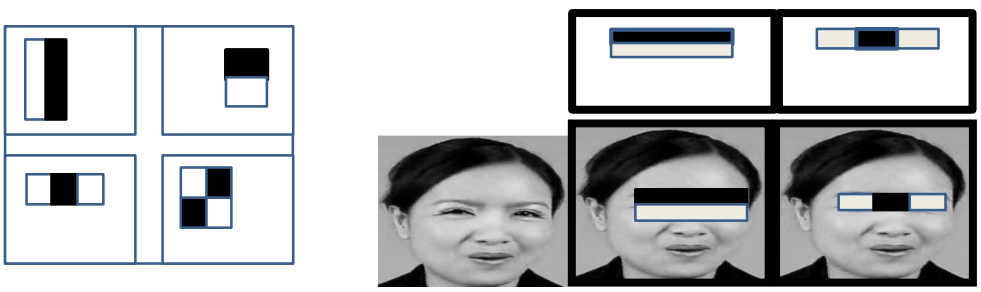


Figure 7. Feature extraction by using Adaboost [5]

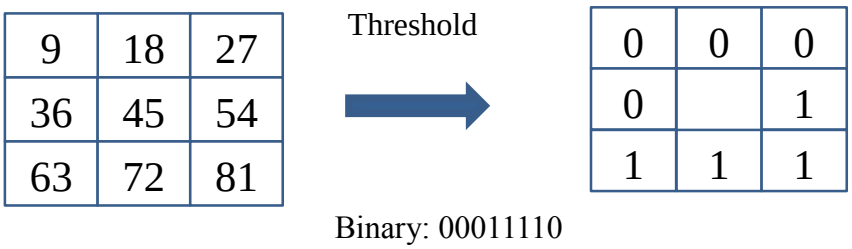


Figure 8. Feature extraction by using LBP [12]

2.3.2 Color Based Method

In facial features extraction, the more commonly used color models are RGB (red, green and blue), HSV (hue, saturation and value) and YCbCr (luminance and chrominance) [13]. RGB colors are primary colors. RGB colors represent HSV and YCbCr colors by varying their color combinations. Firstly, this method transforms RGB colors of the detected face region into YCbCr colors. This colorspace transformation increases the separability between skin and non-skin classes. In YCbCr region obtained after transforming the colorspace, the chrominance information of that region cannot differentiate the features in skin color. The features in the brightness of the color characterize the fairness or darkness of the skin. Therefore, the luminance information extracts the features in skin color. To extract the features of the skin color in the luminance information, the grayscale region from the YCbCr region is converted. In order to obtain the black and white region, the grayscale region is transformed into the binary region with threshold because the features of the skin color are darker than the features of the background color. After thresholding, the morphological operations are applied to remove noise. Then, the facial features are extracted from the binary region. To form a feature vector, these facial features are combined. This feature vector classify the facial expressions such as question, negation, neutral, happy, sad, angry, fear, surprise and so on.

2.3.3 Texture Based Method

For texture features extraction, two-dimensional (2-D) Gabor filters from the 2-D Gabor functions are built. The convolution kernel of a 2-D Gabor function is a product of Gaussian and cosine function. The following equation implements the convolutions of an input image with a 2-D Gabor function:

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \quad (2)$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

where wavelength (λ) specifies its value $\lambda = 2$ or $\lambda > 2$ in pixels. The orientation (θ) specifies its value between 0 and 360 in degrees. The phase offset (φ) specifies its value between -180 and 180 in degrees. The aspect ratio (γ) specifies the ellipticity of the support of the Gabor function. The bandwidth (b) relates to the ratio ($\frac{\sigma}{\lambda}$) , where σ and λ are the standard deviation of the Gaussian factor of the function and the preferred wavelength. The parameters σ and λ can connect with each other by using the equation $\sigma = 0.56\lambda$.

The preferred orientation and spatial frequency in the 2-D Gabor function characterize the 2-D Gabor filter [21]. The Gabor features are obtained by filtering the detected face region with a set of the 2-D Gabor filters and then these features are combined to form a feature vector. This feature vector classify the facial expressions such as question, negation, neutral, happy, sad, angry, fear surprise and so on.

2.4 Facial Expression Recognition

Over the last decade, facial expression recognition has become an active research area that finds the applications such as human–computer interfaces and human emotion analysis. In human emotion analysis, an automated facial expression recognition system is capable of recognizing the facial changes caused by facial expressions. Therefore, the machine can analyse and interpret the various stages of the development of a human emotion. The facial changes provide critical information about human emotions that directly link to the facial expressions.

Pantic and Rothkrantz [16] proposed an expert system called Integrated System for Facial Expression Recognition (ISFER) for the recognition and emotional classification of facial expressions. There are two stages in their system. The first stage is the ISFER Workbench. By applying multiple feature detection techniques in parallel, the ISFER Workbench represents a framework for hybrid facial feature detection. The second stage is HERCULES to convert low level face geometry into high level facial actions. Figure 9 shows the ISFER structure.

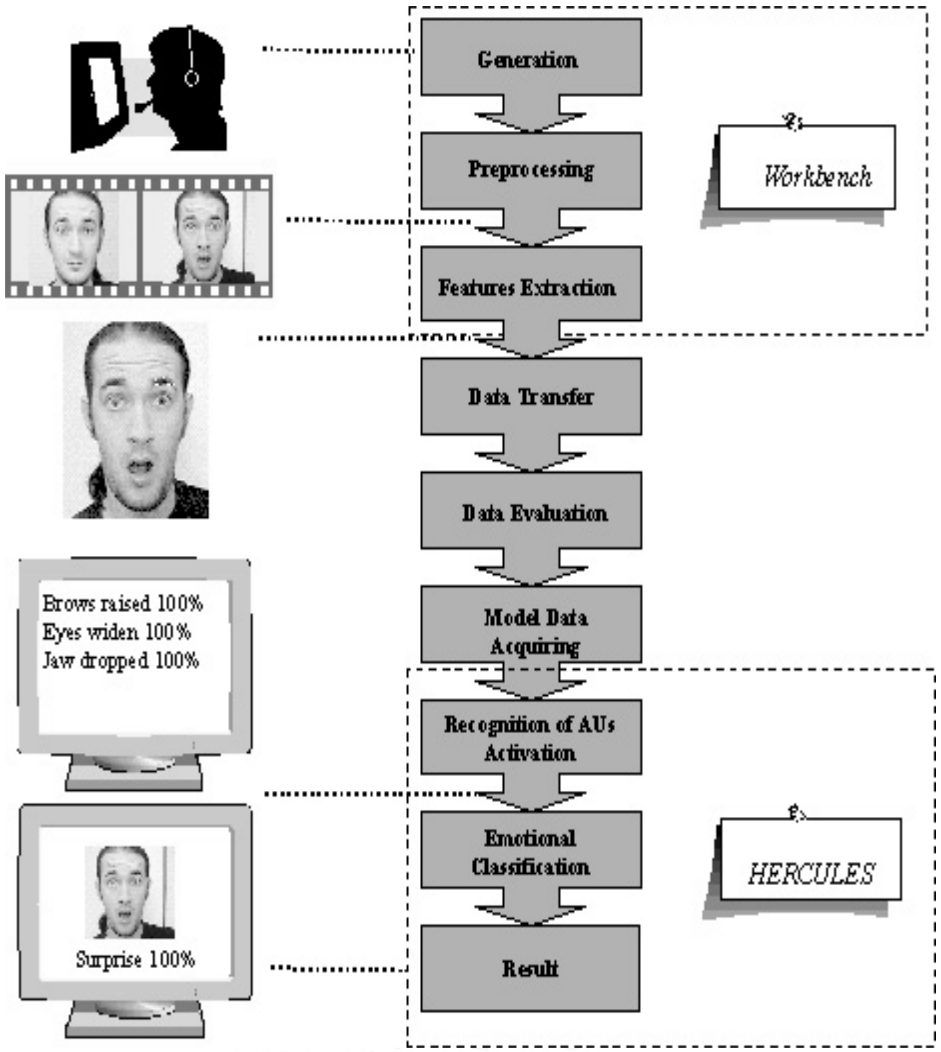


Figure 9. ISFER structure [16]

Their system used a point-based model that made up of the frontal view and side view. The frontal view face model is made up of thirty features. These features were divided into two groups: first group that composed of twenty five features defined in correspondence with a set of nineteen facial points and second group that made up of five features represented specific shapes for the mouth and chin. The points of first group were demonstrated in figure 10.

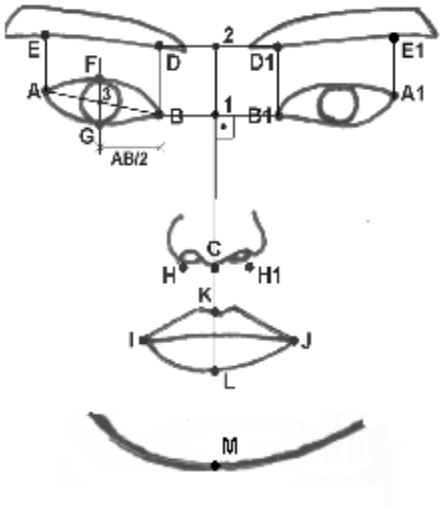


Figure 10. Facial points of the frontal-view [16]

The side view face model was made up of 10 profile points that correspond with the peaks and valleys of the curvature of the profile contour function. The profile points were demonstrated in figure 11.

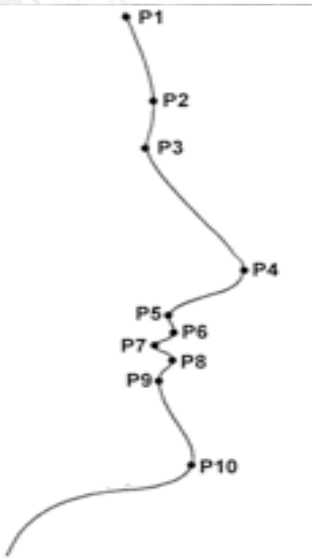


Figure 11. Face profile points [16]

In their system, they utilized multiple feature detection techniques to localize the contours of the prominent facial features such as eyebrows, eyes, nose, mouth, and profile. Then, the model features are extracted from an input dual-view. To select the best of the acquired (redundant) results, their system based on the confidence in a specific detector performance and the knowledge of the facial anatomy. The knowledge of the facial anatomy is used to check the correctness of a certain detector result. To

test the performance of the detection scheme, they used 496 dual views. Their system can deal with face images without facial hair or glasses and cannot deal with minor inaccuracies of the extracted facial features.

In their method, the multidetector processing automatically detects the facial features of the examined facial image. After detecting the contours of the facial features, the system extracts the model features. Then, they calculate the difference between the same detected features in an expressionless face and the currently detected features of the same person. By the production rules, the calculated model deformation is classified as the appropriate AUs-classes based on the knowledge acquired from FACS [17]. For the testing performance of the system, they used 496 dual views. For the lower face AUs, the average recognition rate was 86%. For the upper face AUs, the average classification rate was 92%.

To recognize the facial expression, Yoneyama et al. [18] presented discrete type Hopfield neural networks that are capable to recognize the preliminarily specific facial expressions. Figure 12 shows the flow chart of pattern recalling process of Hopfield neural networks.

For face detection, Yoneyama et al. [18] utilized an analytic approach. In their system, they automatically extracted the height of the mouth, the height of the eyes, and the outer corners of the eyes. After extracting these features, these facial regions are normalized and resized into 8x10 pixels.

For face representation, they used a hybrid approach. By using the optical flow method presented by Horn and Schunck [19], they calculated the optical flow between an extracted facial expression image and a neutral. After calculating the optical flow, the magnitude and the direction are simplified as the information about a vertical movement. To classify facial appearance changes included a horizontal movement, the method will be fail. Their system can correctly encounter the faces without no rigid head motion, glasses and facial hair.

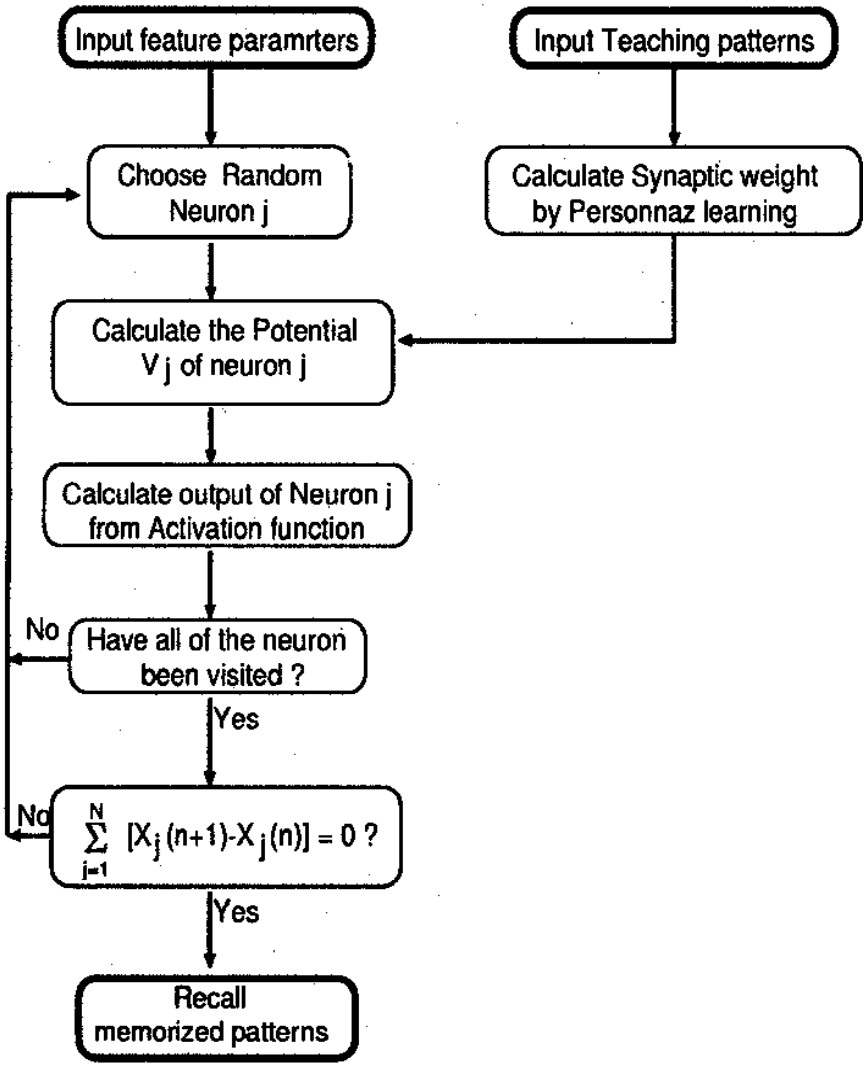


Figure 12. Flow chart of pattern recalling process of Hopfield neural networks [18]

In their system, 80 facial movement parameters are extracted. These parameters described the difference between an expressionless face and the currently extracted facial expression of the same person. To recognize the facial expressions, they use the ranging between 1 and -1 and two identical discrete Hopfield networks. By using the Personnaz learning rule [20], two identical discrete Hopfield networks were trained on the 40 data represented the facial expressions and 4 data represented the most clearly shown facial expressions. In the training of the NN1, all of the examples match the output of the NN1 by using Euclidean distances. In order to decide a category of the facial expressions, the difference among the first minimal average and the second minimal average is greater than 1. Otherwise, the examples in the training of the NN2

match the output of the NN2 according to decide a final category of the facial expressions. By using images in the training of the networks as their testing images, the average classification rate of their system was 92 %.

CHAPTER III

PROPOSED METHOD

3.1 Proposed System

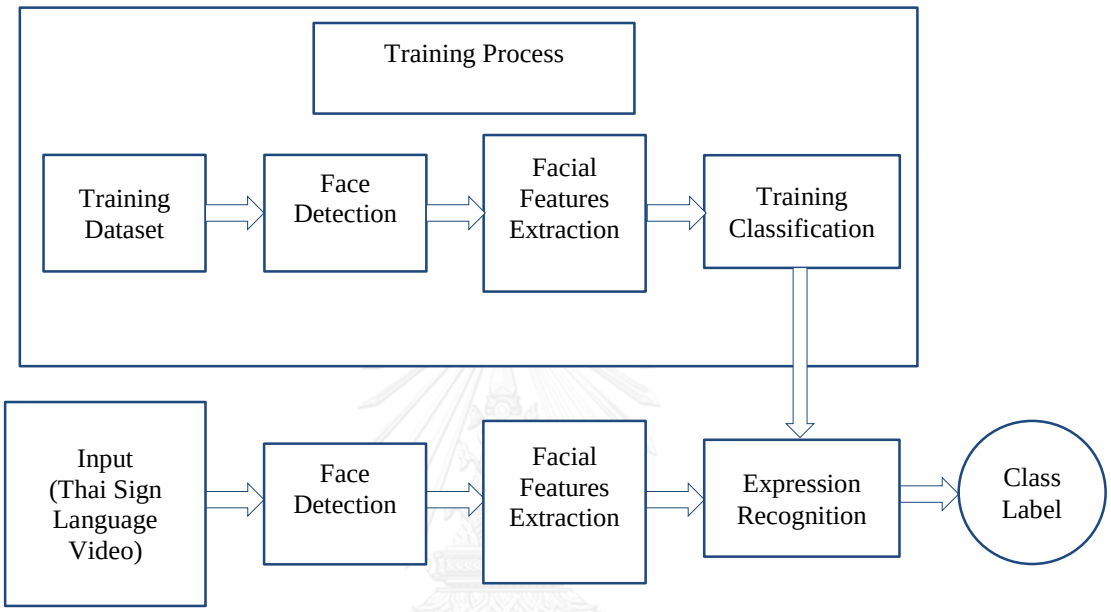


Figure 13. Block Diagram of Proposed System

In this proposed framework, the algorithm of facial expression recognition composes of two main processes: training process and testing process. Each process consists of four steps (figure 9): dataset, face detection, facial features extraction and expression recognition. For training and testing processes, each dataset includes the images selected from Thai sign language videos. These images relate to emotion of the signer.

Face detection is the second stage in this proposed framework. In face detection process, Viola-Jones face detector automatically detects the face regions for the input images of the Thai sign language videos. For Viola-Jones face detection, the detail introduces in Section 3.2.

After detecting the face regions, the next stage is facial features extraction. In this proposed framework, the detected facial regions are normalized to extract the important features. After obtained the normalized features, these features combine to

form a feature vector used in the recognition. Section 3.3 discusses the details about the facial feature extraction.

For the recognition in the last stage, a machine learning technique is required to classify the facial expressions. After extracting the facial features, these features are trained and tested by the machine learning technique to build a classifier. After the recognition module, the output is expressed as class label.

3.2 Face Detection

Face detection is an important part of facial expression recognition because it is the first step of automatic facial expression recognition. Face detection is difficult because it has the variations of the image appearance such as illumination, occlusion, pose variation (frontal, non-frontal) and image orientation. For face detection, the input images are acquired from the Thai sign language videos. In Thai sign language videos, the signer's face expresses the sign with frontal and near-frontal views of face. Therefore, Viola-Jones algorithm is used to detect the face regions in this proposed framework. This algorithm automatically detects the frontal face with high detection rate. All of the detected face regions are shown in figure 10.

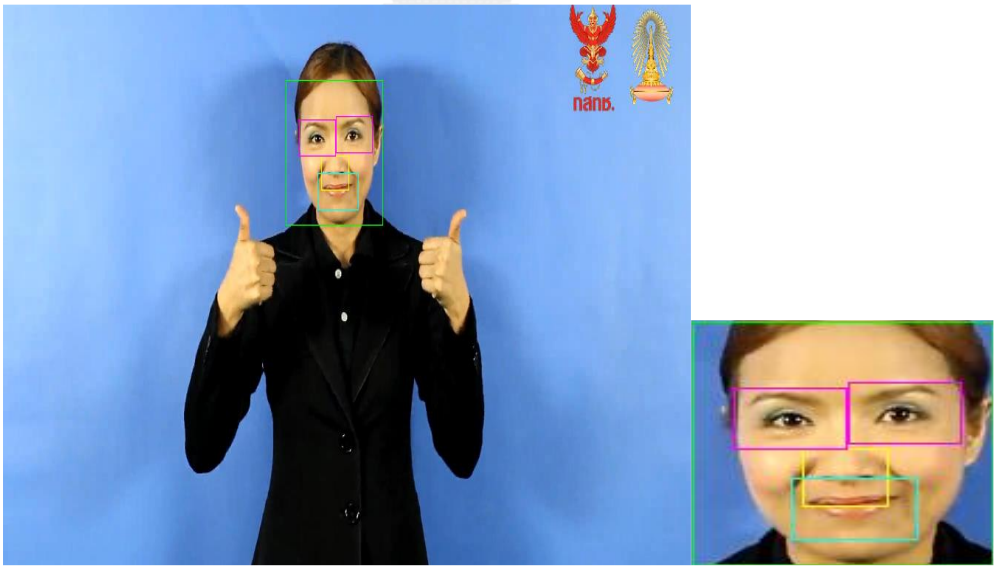


Figure 14. The detected face components image

In Viola-Jones face detection [4]-[5], all Haar-like features consist of three basic types of rectangle features in figure 11: two-rectangle, three-rectangle and four-rectangle features. The value of a two-rectangle feature is the difference of the summed pixels between black and white rectangles. A three-rectangle feature calculates the

difference between the sum within two outside white rectangles and the sum in a centre black rectangle. Finally a four-rectangle feature computes the difference between the sums of two rectangle pairs in the diagonal.

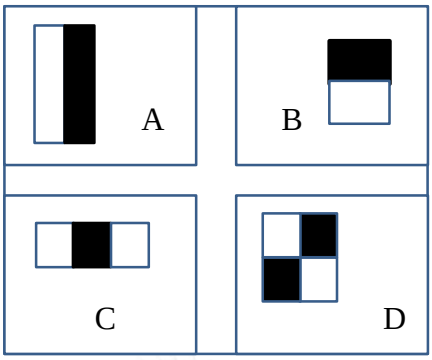


Figure 15. Basic Haar-like feature types [5]

A feature is evaluated by the difference between the summed pixel values of black rectangle and that of white rectangle.

Rectangle features:

$$f(x) = \sum(\text{pixels in black area}) - \sum(\text{pixels in white area}) \qquad (3)$$

The features locate at any area within a scanning image window. A 24x24 window results over 160000 features. The number of rectangle features is large. Since evaluating the entire set of features is still extremely expensive, it cannot be performed by a real-time application. A very small number of the extracted features can be used to form an effective classifier for face detection. AdaBoost [6] is a machine learning algorithm which selects only the best features among all these rectangle features and trains the classifier.

These rectangle features (also known as weak classifiers) cannot classify the image. However, these rectangle features form a strong classifier by combining with other rectangle features. As a linear combination of weighted simple weak classifiers, Adaboost builds a strong classifier.

The cascade classifier is made up of stages. Each stage is composed of an AdaBoost strong classifier. Each stage is used to determine if a sub-window contains a face or non-face. If a sub-window fails in the first stage, it is rejected as non-face. If a

sub-window passes in the first stage, it is applied to the second stage as face and the process is continued by adding stages until the target detection is met.

3.3 Facial Features Extraction

Facial features extraction is one of the most important steps for the facial expression recognition because facial features can make a better recognition of the facial expression. After detection process, we extract all of the face regions. In the facial features extraction, we extract the features by averaging and normalizing the face regions.

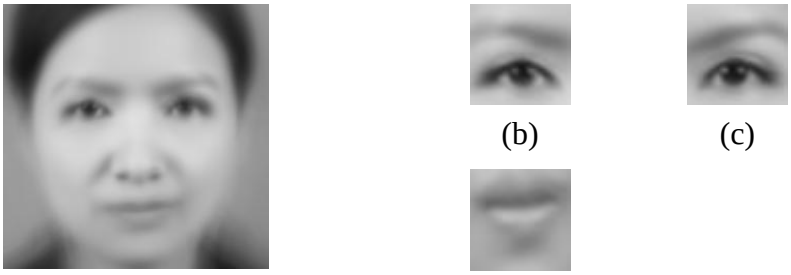
To improve the performance of the proposed system, all of the extracted face regions are resized to a uniform size. Then, all of the resized face regions are converted into grayscale images. After converting grayscale images, the following equation averages and normalizes all of the face regions to remove the variation effects in the scale of the images,

$$x_{normalized} = \frac{|x_i - \mu|}{\sigma} \tag{4}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{6}$$

where x_i is the individual pixel values of images, n is the total number of images, μ is the mean averaging and σ^2 is the variance of images. After obtaining the normalized features, we combine these features to form a feature vector. This feature vector is used to classify the facial expressions such as positive, negative and neutral. In figure 12, we show the results of mean vector.



(a) (b) (c) (d)

Figure 16. Results of mean vector (a) Face image (b) Right eye image (c) Left eye image and (d) Mouth image

3.4 Facial Expression Recognition

3.4.1 K-Nearest Neighbors (KNN)

After facial features extraction process, we calculate the classification accuracy of the face regions by using KNN algorithm. KNN is a simple algorithm for facial expression recognition. In training, the images are averaged and normalized in a given class to generate a template for this class. After training, the input images are matched with the closest template by using distance metrics. These distance metrics measure the dissimilarity between the input image and the closest template. In our proposed algorithm, we use four distance metrics: 'Euclidean', 'City block', 'Cosine' and 'Correlation' to calculate the classification accuracy in the KNN algorithm.

The Euclidean distance between two points, a and b, with k dimensions is calculated as:

$$d = \sqrt{\sum_{j=1}^{k} (a_j - b_j)^2} \tag{7}$$

Euclidean distance between two data points involves computing the square root of the sum of the squares of the differences between corresponding values.

The City block distance between two points, a and b, with k dimensions is calculated as:

$$d = \sum_{j=1}^{k} \left| a_j - b_j \right| \tag{8}$$

The Cosine correlation between two points, a and b, with k dimensions is calculated as:

$$d = \frac{\sum_{j=1}^{k} a_j \times b_j}{norm(a) \times norm(b)} \tag{9}$$

where

$$norm(a) = \sqrt{\sum_{j=1}^{k} a_{j^2}} \tag{10}$$

$$norm(b) = \sqrt{\sum_{j=1}^{k} b_{j^2}} \tag{11}$$

The Correlation between two points, a and b, with k dimensions is calculated as:

$$d = \frac{\text{cov}(a,b)}{std(a) \times std(b)} \tag{12}$$

where

$$\text{cov}(a,b) = \frac{1}{k}\sum_{j=1}^{k}(a_j - \bar{a}) \times (b_j - \bar{b}) \tag{13}$$

$$std(a) = \sqrt{\frac{1}{k}\sum_{j=1}^{k}(a_j - \bar{a})^2} \tag{14}$$

$$\bar{a} = \frac{1}{k}\sum_{j=1}^{k} a_j \tag{15}$$

3.4.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a popular technique to classify the facial expressions. By using a hyperplane, SVM divides the space into two regions. Each region classifies the one type of element. In training, a margin separator is defined from the negative training samples and positive training samples by using mathematical functions. In testing, the testing sample is considered as positive case or negative case.

Given a training set of labelled examples $T = \{(x_i, y_i), i = 1,\ldots l\}$ where $x_i \in R_n$ and $y_i \in \{1, -1\}$, the following equation classifies the new test data x:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \quad (16)$$

where α_i are Lagrange multipliers of a dual optimization problem, and $K(x_i,x)$ is a kernel function. Given a nonlinear mapping $\emptyset$ that embeds input data into feature space, kernels have the form of $K(x_i,x_j) = < \emptyset(xi) . \emptyset(x_j) >$. To separate the training data in feature space, SVM finds a linear separating hyperplane with the maximal margin. The parameter b is the optimal hyperplane. Estimating the optimal hyperplane is equivalent to solving a linearly constrained quadratic programming problem.

CHAPTER IV

Experiments and Results

The facial changes are the responses of the signer's emotional states. Therefore, we can recognize the signer's emotion by facial changes such as raising of the eyebrows in surprise. Facial expressions still have challenges because some facial changes are similar to identify facial expressions. For example, we understand surprise expression by opening of the mouth and raising of the eyebrows. By pulling of the mouth corner and raising of the eyebrows, we describe happy expression. Raising of the eyebrows is similar in surprise and happy expressions. This similarity of facial changes can reduce the accuracy of the facial expression recognition. Therefore, we need to advance our understanding of each facial expression and discover new ways of improving facial expression recognition.

4.1 Dataset

In this proposed framework, positive, negative and neutral expressions are recognized to advance our understanding of each facial expression. To improve the accuracy of the facial expression recognition, the dissimilarity of facial changes is used for positive, negative and neutral expressions. Positive expression is identified by the facial changes: raising of the eyebrows and pulling of the mouth corner. Lowering of the eyebrows and puckering of the mouth are chosen to identify negative expression. For neutral expression, expressionless facial changes are selected.

For the dataset, we select 360 image sequences from the Thai sign language videos. Then, we separate the dataset into three groups: positive dataset, negative dataset and neutral dataset. Positive dataset is the collection of good sentimental images such as happy, delicious, understand and so on. Figure 13 shows the examples of positive dataset. Negative dataset is the collection of bad sentimental images such as punch, be lost, danger and so on. The examples of negative dataset are shown in figure 14. Neutral dataset is the collection of images that have no significant sentimental expressions such as bank, box, clock and so on. Figure 15 shows the examples of neutral dataset.

For the training dataset, we choose positive expression images (100), negative expression images (100) and neutral expression images (100). Similarly, we select positive expression images (20), negative expression images (20) and neutral expression images (20) as the testing dataset. For our experiment, Viola-Jones algorithm automatically detects the faces from these selected image sequences. After face detection, we extract the face, mouth, right eye and left eye regions. The extracted face, eyes including eyebrows and mouth are resized into 256 x 256 pixels, 50 x 50 pixels and 70 x 40 pixels.

face

mouth

right eye

left eye

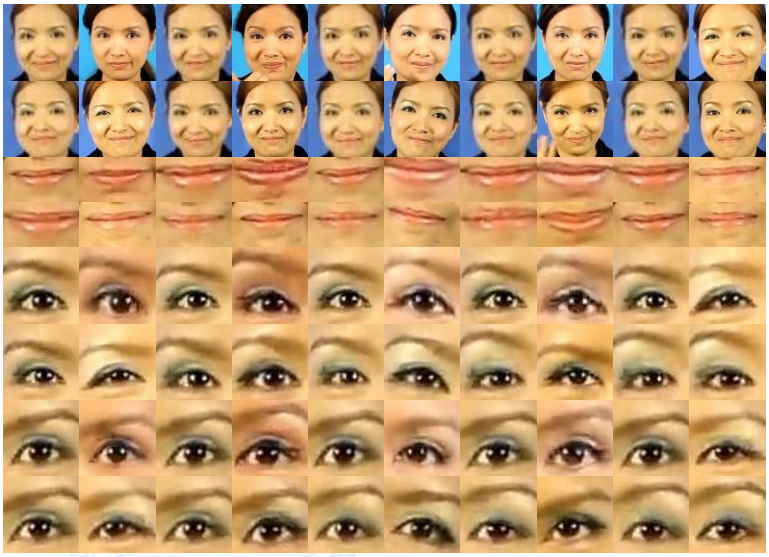


Figure 17. Examples of positive dataset

face

mouth

right eye

left eye

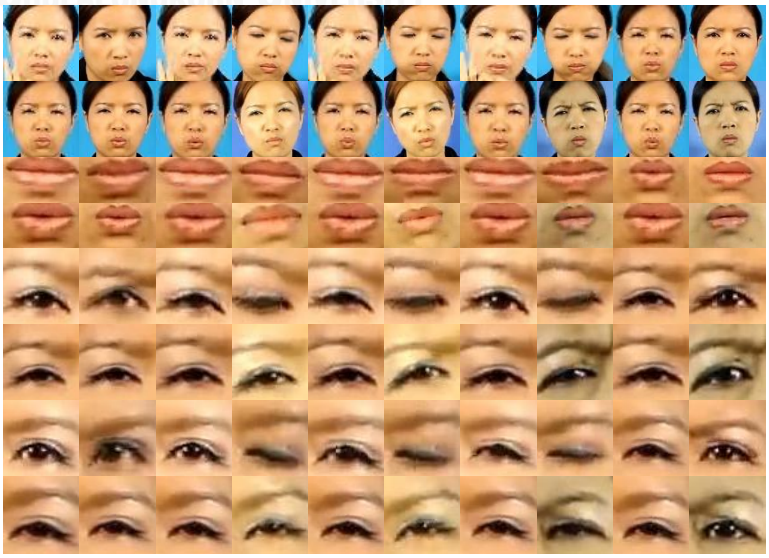


Figure 18. Examples of negative dataset

Figure 19. Examples of neutral dataset

In our proposed algorithm, the classification accuracy is compared by recognizing each part of the face and the combination of each part of the face. For the recognition process, KNN and SVM algorithms are used to calculate the classification accuracy. In the KNN algorithm, there are four distance metrics: 'Euclidean', 'City block', 'Cosine' and 'Correlation'. These distance metrics are used to compute the classification accuracy in the KNN algorithm.

4.2 Testing Dataset and Training Dataset

Table 1. Classification accuracy by using KNN with Euclidean

Class	Euclidean
Face	73.33
Mouth	**85**
Right eye	70
Left eye	53.33
Right eye and left eye	68.33
Right eye, left eye and mouth	**78.33**
Right eye, left eye, mouth and face	76.67

After calculating the results shown in Table 1, it is found that the accuracy rate of the mouth is better than the accuracy rates of another each part of the face. The classification accuracy of the mouth is 85%. This accuracy is the highest classification accuracy because mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the accuracy of the combination of the eyes and mouth is 78.33%. This accuracy is higher than another classification accuracy of the combination of each part of the face.

Table 2. Classification accuracy by using KNN with City block

Class	City block
Face	68.33
Mouth	**85**
Right eye	61.67
Left eye	56.67
Right eye and left eye	68.33
Right eye, left eye and mouth	**75**
Right eye, left eye, mouth and face	73.33

In Table 2, the experimental results show that the recognition rate of the mouth is 85%. This recognition rate is the highest rate in each part of the face because mouth is more dissimilar than facial changes of the eyes. The recognition rate of the combination of the eyes and mouth is 75% in the combination of each part of the face. This recognition rate is higher in comparison to the recognition rates of another combination of each part of the face.

Table 3. Classification accuracy by using KNN with Cosine

Class	Cosine
Face	73.33
Mouth	**90**
Right eye	63.33
Left eye	70
Right eye and left eye	68.33
Right eye, left eye and mouth	**80**
Right eye, left eye, mouth and face	73.33

Table 3 shows that the recognition performance of the mouth is 90%. This recognition performance is the highest performance in each part of the face. The reason is that mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the recognition performance of the combination of the eyes and mouth is 80%. This recognition performance is higher than the recognition performances of another combination of each part of the face.

Table 4. Classification accuracy by using KNN with Correlation

Class	Correlation
Face	73.33
Mouth	**85**
Right eye	61.67
Left eye	56.67
Right eye and left eye	63.33
Right eye, left eye and mouth	**81.67**
Right eye, left eye, mouth and face	75

After computing the results shown in Table 4, it is found that the accuracy rate of the mouth is 85%. This accuracy rate is the highest rate in each part of the face due

to mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the accuracy rate of the combination of the eyes and mouth is 81.67%. This accuracy is better than the accuracy rates of another combination of each part of the face.

Table 5. Classification accuracy by using SVM

Class	SVM
Face	61.67
Mouth	**88.33**
Right eye	68.33
Left eye	63.33
Right eye and left eye	60.00
Right eye, left eye and mouth	**83.33**
Right eye, left eye, mouth and face	73.33

In Table 5, the experimental results show that the recognition rate of the mouth is 88.33%. This recognition rate is the highest rate in each part of the face because mouth is more dissimilar than facial changes of the eyes. The recognition rate of the combination of the eyes and mouth is 83.33% in the combination of each part of the face. This recognition rate is higher in comparison to the recognition rates of another combination of each part of the face.

4.6 Testing dataset with Gaussian blur and training dataset

Table 6. Classification accuracy by using KNN with Euclidean

Class	Euclidean
Face	73.33
Mouth	**81.67**
Right eye	60
Left eye	53.33
Right eye and left eye	56.67
Right eye, left eye and mouth	73.33
Right eye, left eye, mouth and face	**75**

After computing the results shown in Table 6, it is found that the accuracy rate of the mouth is 81.67%. This accuracy rate is the highest rate in each part of the face due to mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the accuracy rate of the combination of the eyes, mouth and face is 75%. This accuracy is better than the accuracy rates of another combination of each part of the face.

Table 7. Classification accuracy by using KNN with City block

Class	City block
Face	68.33
Mouth	**78.33**
Right eye	60
Left eye	56.67
Right eye and left eye	61.67
Right eye, left eye and mouth	**75**
Right eye, left eye, mouth and face	**75**

Table 7 shows that the recognition performance of the mouth is 78.33%. This recognition performance is the highest performance in each part of the face. The reason is that mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the recognition performance of the combination of the eyes is 61.67%. This recognition performance is lower than the recognition performances of another combination of each part of the face.

Table 8. Classification accuracy by using KNN with Cosine

Class	Cosine
Face	73.33
Mouth	**86.67**
Right eye	61.67
Left eye	51.67
Right eye and left eye	46.67
Right eye, left eye and mouth	71.67
Right eye, left eye, mouth and face	**73.33**

In Table 8, the experimental results show that the recognition rate of the mouth is 86.67%. This recognition rate is the highest rate in each part of the face because mouth is more dissimilar than facial changes of the eyes. The recognition rate of the combination of the eyes, mouth and face is 73.33% in the combination of each part of the face. This recognition rate is higher than the recognition rates of the combination of each part of the face.

Table 9. Classification accuracy by using KNN with Correlation

Class	Correlation
Face	73.33
Mouth	**78.33**
Right eye	58.33
Left eye	56.67
Right eye and left eye	58.33
Right eye, left eye and mouth	**81.67**
Right eye, left eye, mouth and face	75

After calculating the results shown in Table 9, it is found that the classification accuracy of the mouth is 78.33%. This classification accuracy is better than the classification accuracies of another each part of the face. The reason is that mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the accuracy of the combination of the eyes and mouth is 81.67%. This accuracy is higher than another classification accuracy of the combination of each part of the face.

Table 10. Classification accuracy by using SVM

Class	SVM
Face	61.67
Mouth	**90**
Right eye	50
Left eye	58.33
Right eye and left eye	45
Right eye, left eye and mouth	**80**
Right eye, left eye, mouth and face	73.33

Table 10 shows that the recognition performance of the mouth is 90%. This recognition performance is the highest performance in each part of the face. The reason is that mouth is more dissimilar than facial changes of the eyes. For the combination of each part of the face, the recognition performance of the combination of the eyes and mouth is 80%. This recognition performance is higher than the recognition performances of another combination of each part of the face.

Table 11. Averaged classification accuracy for testing dataset and training dataset

Class	Euclidean	City block	Cosine	Correlation	SVM
Face	73.33	68.33	73.33	73.33	61.67
Mouth	**85**	**85**	**90**	**85**	**88.33**
Right eye	70	61.67	63.33	61.67	68.33
Left eye	53.33	56.67	70	56.67	63.33
Right eye and left eye	68.33	68.33	68.33	63.33	60.00
Right eye, left eye and mouth	**78.33**	**75**	**80**	**81.67**	**83.33**
Right eye, left eye, mouth and face	76.67	73.33	73.33	75	73.33

For testing dataset and training dataset in Table 11, the experimental results show that the recognition rate of the mouth is higher than another recognition rate of each part of the face. The reason is that mouth is more dissimilar than facial changes of the eyes. The eyes get lower accuracy rate than the accuracy rate of the mouth due to similar facial changes. For the combination of each part of the face, the classification accuracy of the combination of the eyes and mouth is better in comparison to the classification accuracies of another combination of each part of the face.

Table 12. Averaged classification accuracy for testing dataset with Gaussian blur and training dataset

Class	Euclidean	City block	Cosine	Correlation	SVM
Face	73.33	68.33	73.33	73.33	61.67
Mouth	**81.67**	**78.33**	**86.67**	**78.33**	**90**
Right eye	60	60	61.67	58.33	50
Left eye	53.33	56.67	51.67	56.67	58.33
Right eye and left eye	56.67	61.67	46.67	58.33	45
Right eye, left eye and mouth	73.33	**75**	71.67	**81.67**	**80**
Right eye, left eye, mouth and face	**75**	**75**	**73.33**	75	73.33

Table 12 shows the recognition performances for testing dataset with Gaussian blur and training dataset. The experiments show that the accuracy rate of the mouth is higher in comparison to another accuracy rate of each part of the face. The classification accuracy of the eyes is lower than the classification accuracy of the mouth. The reason is that mouth is more dissimilar than facial changes of the eyes. The classification accuracy of the combination of the eyes and mouth is better than another classification accuracy of the combination of each part of the face.

CHAPTER V

Conclusions and Future Works

5.1 Conclusions

For the facial expressions recognition, our algorithm applies two experiments: testing dataset and testing dataset with Gaussian blur. In our proposed algorithm, we compare the classification accuracy for each part of the face and the combination of each part of the face by using KNN and SVM algorithms. The results show that the recognition rate of the mouth is higher than the classification accuracies of another part of the face because mouth is more dissimilar than facial changes of the eyes. The recognition rate of the combination of the eyes and mouth is better in comparison to the recognition rates of another combination of each part of the face. The reason is that the similar facial changes reduce the classification accuracy rate.

5.2 Future Works

In this proposed algorithm, there are still limitations which need to be addressed. This proposed facial expression recognition has been designed for frontal views of face images. By further extension, the face detection technique for the multiple views of face images could be implemented for facial expression recognition. The more the input images in the training dataset, the more the performance of facial expression recognition. Moreover, the classification accuracy of facial expression recognition could be improved by using other algorithms and methods.

References

[1] S. Aramvith, T. Chauksuvanit, “ELECTRONICS THAI SIGN LANGUAGE (e-TSL) COMMUNICATION SYSTEM,” In ITU-T Focus Group Audio Visual Media Accessibility, 2012.

[2] Saengsri, Supawadee, Vit Niennattrakul, and Chotirat Ann Ratanamahatana. "TFRS: Thai finger-spelling sign language recognition system." Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on. IEEE, 2012.

[3] T. Suksil, T. H. Chalidabhongse, “Hand Detection and Feature Extraction for Thai Sign Language Recognition,” In International Workshop on Advanced Image Technology, 2012.

[4] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[5] Viola, Paul, and Michael Jones. "Robust real-time object detection."International Journal of Computer Vision 4 (2001): 51-52.

[6] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." Journal of computer and system sciences 55.1 (1997): 119-139.

[7] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003).

[8] Oz, Cemil, and Ming C. Leu. "American Sign Language word recognition with a sensory glove using artificial neural networks." Engineering Applications of Artificial Intelligence 24.7 (2011): 1204-1213.

[9] Liwicki, Stephan, and Mark Everingham. "Automatic recognition of fingerspelled words in British sign language." Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, 2009.

[10] Woodward, James. "Modern standard Thai Sign Language, influence from ASL, and its relationship to original Thai sign varieties." Sign Language Studies 92.1 (1996): 227-252.

[11] Holden, Eun-Jung, Gareth Lee, and Robyn Owens. "Australian sign language recognition." Machine Vision and Applications 16.5 (2005): 312-320.

[12] Liao, Shu, et al. "Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features." Image Processing, 2006 IEEE International Conference on. IEEE, 2006.

[13] Sobottka, Karin, and Ioannis Pitas. "Face localization and facial feature extraction based on shape and color information." Image Processing, 1996. Proceedings., International Conference on. Vol. 3. IEEE, 1996.

[14] Rowley, Henry, Shumeet Baluja, and Takeo Kanade. "Neural network-based face detection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 20.1 (1998): 23-38.

[15] Heisele, Bernd, et al. "Component-based face detection." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.

[16] Pantic, Maja, and Leon JM Rothkrantz. "Expert system for automatic analysis of facial expressions." Image and Vision Computing 18.11 (2000): 881-905.

[17] Ekman, Paul, and Erika L. Rosenberg. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, 1997.

[18] Yoneyama, Masahide, et al. "Facial expressions recognition using discrete hopfield neural networks." Image Processing, 1997. Proceedings., International Conference on. Vol. 1. IEEE, 1997.

[19] Horn, Berthold K., and Brian G. Schunck. "Determining optical flow." 1981 Technical symposium east. International Society for Optics and Photonics, 1981.

[20] Kanter, I., and Haim Sompolinsky. "Associative recall of memory without errors." Physical Review A 35.1 (1987): 380.

[21] Jemaa, Yousra Ben, and Sana Khanfir. "Automatic local Gabor features extraction for face recognition." arXiv preprint arXiv:0907.4984 (2009).

[22] Soontranon, N., Supavadee Aramvith, and Thanarat H. Chalidabhongse. "Face and hands localization and tracking for sign language recognition."Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on. Vol. 2. IEEE, 2004.

REFERENCES

APPENDIX

VITA

May Thandar Htay was born in Magway, Myanmar, in 1987. She obtained bachelor degree from Department of Electronics Communication, Magway Technological University, Myanmar on 2008. She has been granted a scholarship by the AUN/SEED-Net (www.seed-net.org) to pursue her Master degree in electrical engineering at Chulalongkorn University, Thailand, since 2013. She conducted her graduate study with the Digital Signal Processing Research Laboratory, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University. Her research interest includes Multimedia and Video Processing, and Computer Vision for surveillance system.