ระบบรู้จำฟันบนพื้นฐานของโครงข่ายประสาทโดยใช้ลักษณะเด่นผสม

นางสาวสุปรัชญา วีระประสิทธิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553

Neural Network-based Teeth Recognition System using Hybrid Features

Miss Suprachaya Veeraprasit

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science and Information

Department of Mathematics

Faculty of Science

Chulalongkorn University

Academic Year 2010

Thesis Title	NEURAL NETWORK-BASED TEETH RECOGNITION SYSTEM USING HYBRID FEATURES
By	Miss SUPRACHAYA VEERAPRASIT
Field of Study	Computer Science and Information
Thesis Advisor	Suphakant Phimoltares, Ph.D.

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

..... Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

..... Chairman

(Assistant Professor Saranya Maneeroj, Ph.D.)

..... Thesis Advisor

(Suphakant Phimoltares, Ph.D.)

..... External Examiner

(Atchara Mahaweerawat, Ph.D.)

สุปรัชญา วีระประสิทธิ์ : ระบบรู้จำฟันบนพื้นฐานของโครงข่ายประสาทโดยใช้ลักษณะเด่นผสม. (NEURAL NETWORK-BASED TEETH RECOGNITION SYSTEM USING HYBRID FEATURES) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : อ.ดร. ศุภกานต์ พิมลธเรศ 52 หน้า.

ในปัจจุบันเทคโนโลยีที่ใช้ลักษณะทางชีวภาพของบุคคลเพื่อยืนยันตัวบุคคลนั้นได้ใช้อย่างแพร่หลายในระบบรักษาความปลอดภัยต่างๆประสิทธิภาพของระบบปฏิบัติ-การเหล่านั้นขึ้นอยู่กับประเภทของข้อมูลลักษณะทางชีวภาพที่นำมาใช้ยืนยันตัวบุคคล อย่างไรก็ตามข้อมูลบางชนิดสามารถปลอมแปลงได้โดยการทำศัลยกรรมโดยเจตนาหรือบุคคลเหล่านั้นถูกเปลี่ยนแปลงโดยไม่ได้ตั้งใจ เช่น ใบหน้า ม่านตา รอยพิมพ์ฝ่ามือ และ รอยพิมพ์นิ้วมือ เมื่อเทียบกับสิ่งเหล่านี้ ฟันจัดเป็นอวัยวะที่ไม่สามารถถูกแปลงรูปร่างได้โดยง่าย ในวิทยานิพนธ์ฉบับนี้ได้นำเสนอลักษณะเด่นที่เหมาะสมพร้อมกับตัวแบบโมเดลการเรียนรู้ของเครื่องสำหรับการรู้จำฟัน ลักษณะเด่นผสมของระบบนี้ประกอบด้วยลักษณะเด่นทั่วไปและลักษณะเด่นเฉพาะซึ่งถูกส่งเข้าไปในระบบพร้อมกัน ในวิทยานิพนธ์ฉบับนี้ ลักษณะเด่นทั่วไปที่ถูกนำเสนอได้ทำการวิเคราะห์จนได้ผลการทดลองที่สมควร ลักษณะเด่นเหล่านี้ประกอบด้วยค่าที่ได้จากการแยกค่าเชิงเดี่ยวและ ฮิสโทแกรมสีของรูปฟัน ส่วนลักษณะเด่นเฉพาะที่ได้นำเสนอนั้นคืออัตราส่วนความกว้างของฟันจากฟันหน้าด้านบน ลักษณะเด่นเหล่านี้ได้ถูกส่งเข้าโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบหลายชั้นพร้อมขั้นตอนวิธีการแพร่ย้อนกลับแบบเลเวนเบิร์ก-มาร์ควอทด์ด้วยลักษณะเด่นเหล่านี้ วิธีที่ได้นำเสนอได้แสดงให้เห็นว่าดีกว่าวิธีการที่มีอยู่แล้วในเชิงความแม่นยำและข้อผิดพลาด

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต.....

สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2553

5273865123 : MAJOR COMPUTER SCIENCE AND INFORMATION

KEYWORDS : TEETH IDENTIFICATION / BIOMETRIC / NEURAL NETWORK / HYBRID FEATURES / LEVENBERG-MARQUARDT ALGORITHM

SUPRACHAYA VEERAPRASIT: NEURAL NETWORK-BASED TEETH RECOGNITION SYSTEM USING HYBRID FEATURES. ADVISOR: SUPHAKANT PHIMOLTARES, Ph.D. , 52 pp.

Nowadays, biometric technology is used in various security applications. The efficiency of such applications depends upon a type of biometric information. Nevertheless, some information can be faked by intent surgery or they are unexpectedly reshaped such as face, iris, palmprint and fingerprint. Unlike ordinary features, teeth cannot be easily reshaped. In this thesis, hybrid features and machine learning model for teeth recognition are proposed. Hybrid features of this system are composed of global and local features simultaneously fed into the system. In this thesis, proposed global features composed of singular values from singular value decomposition and color histogram of teeth image are analyzed and give the adequate result whilst the proposed local features are the ratio of the width from upper-front-teeth. These features were fed into the multilayer perceptron network with Levenberg-Marquart backpropagation training algorithm. With these features and model, the proposed method performs better than other existing techniques in terms of accuracy and error.

Department : Mathematics Student's Signature

Field of Study : Computer Science and Information Advisor's Signature

Academic Year : 2010

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. Suphakant Phimoltares, at The Advanced Virtual and Intelligent Computing (AVIC) Research Center, for all his great support helping me to improve my researching skill, and giving me the great opportunity in publishing proceedings.

In addition, the research will not be complete without the main support from the Center of Excellent in Mathematics, CHE, Sri Ayutthaya Rd., Bangkok, 10400, Thailand. The Center of Excellent in Mathematics brings the supports via “Research Assistants scholarship.” I would like to appreciate that great support for necessary tuition fees, and other education costs.

CONTENTS

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

With personal identification, there are biometric features that are used to identify person such as face, iris, palmprint and fingerprint image. In this day and age, some features can be faked by intent surgery or they are unexpectedly reshaped. In other hand, this feature can be used in the forensic science. To identify the clay, in some case that face and fingerprint are unusable, the teeth might be usable instead. For example, a person died in fire accident causing the corpse and features are reshaped so such features cannot be used. Unlike ordinary features, teeth cannot be easily reshaped. Thus, if teeth can be used as the main biometric feature, this problem can be solved. Moreover, according to biometric researchers aim to find the method to identify person with the less time consumption and evidence, teeth can be taken into account as the main biometric feature. For instance, a criminal has escaped from the prison and have the surgery. It is possible that face recognition cannot handle this case. With his teeth image, it can guarantee that he is a criminal and arrested by police at last. Additionally from the advantage of teeth together with the requirement of low storage, this technology can reduce cost for development.

People teeth, among the personal features, had been overlooked. We cannot outwardly distinct the person by teeth pattern. In fact, both pattern and color of people teeth are different according to the personal behavior. However, there is a case that their behaviors are the same; their teeth pattern might be different. In this research, teeth pattern is selected as main feature to distinguish people. The image samples were taken from subjects with the same range of age and same environment. Apart from that, neural network with back propagation learning is selected as a main tool to identify pattern because of the advantage in time and space consumption.

Refer to related works about teeth recognition, Kiattisin et al. proposed improved PCA and LDA-Based personal identification method, which uses PCA as identifier and improved the algorithm for filtering the unclassified images. Besides, PCA is also used in

face recognition as PCA based face recognition and testing criteria proposed by B. Poon et al. and enhanced face recognition through variation of principle component analysis by D.A. Meedeniya et al. Those techniques require complicated preprocessing. The other technique used for teeth recognition is applying Difference Image Entropy (DIE) to identify individual person. For this technique, Kim et al. proposed teeth recognition based on multiple attempts in mobile device, while Joen et al. proposed performance evaluation of DIE-based teeth image recognition system. Their method includes the calculation of average entropy from each image. This causes high complexity. In addition, to improve the accuracy of identification, the mix of two different physical features is proposed by Kim et al. Teeth image and voice are combined for multimodal biometric authentication in mobile environment. Although they achieved high accuracy, many facial features were required for the face recognition process.

In this work, mixed features were proposed along with machine learning model for teeth recognition in terms of accuracy and false acceptance. In our experiment, global features are defined by the mixture of SVD of teeth image and color histogram. The local features are analyzed from teeth width ratio to reduce the error affected from the usage of only global features. These features were fed into the neural network based on Levenberg-Marquart backpropagation training algorithm because this algorithm can handle the complex input within the short learning period. To evaluate overall system including proposed features and neural network for teeth recognition, the proposed system was compared with other existing models for teeth recognition such as k-Nearest Neighbor, Naïve Bayes, and Resilient propagation (R-Prop). In addition, the system with mixed features is also compared with the system with only global features to ensure that using mixed features yields better results.

This thesis is organized as follows. The second chapter explains a feature extraction and classification techniques which are used. The third chapter shows the proposed method. The fourth chapter is about experiment and results. The last two chapters are discussion and conclusion respectively. Some parts of this work contains the details which was already proposed in the proceeding of the 2nd International

Conference on Information Engineering and Computer Science 2010 (ICIECS2010) and the proceeding of the IEEE International Conference on Image Information and Techniques 2011 (IST2011)

1.1. Objectives

The main objectives of this study are the following:

1. A teeth recognition that can be used in various applications.
2. The teeth recognition with high accuracy in database.
3. The complete system available for single image.

1.2. Scope

1. The dataset used in this experiment consists of portrait images with smiling.
2. Motion blur errors from hand swaying are acceptable.
3. The teeth recognition system is designed for single image input.

1.3. Research methodology

In order to achieve the objectives, the following tasks will be stated. The detail of the work in each task is described below.

- Study related literatures: In this task, the similar works are analyzed. Existing works were compared with each others for getting the novel idea which is the proposed method.
- Define problem: According to the previous task, the novel idea was defined then we stated the objective and scope of the work.
- Design the methodology: After the problem was stated, the experimental steps were planned.

- Prepare the database: The pictures for testing were taken from the volunteers who are age in range.
- Do the experiment
- Conclude the results for the manuscript submission and thesis

Table 1.1: Task schedule

No.	Task	Month sequence																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Study related literatures	■	■	■	■														
2	Define problem				■	■													
3	Design the methodology				■														
4	Prepare database					■													
5	Do experiment					■	■	■	■	■	■	■	■	■					
6	Conclude the results													■	■	■	■		
7	Prepare the proceedings								■	■		■	■						
8	Prepare thesis proposal test									■	■								
9	Prepare dissertation.										■	■	■	■	■	■	■	■	■

1.4. Benefits

1. A teeth recognition system that can be used in various applications.
2. A single image teeth recognition system with high performance and easy to implement.

CHAPTER II

LITERATURE REVIEW

The biometric information is used to do the person identification extensively. Using biometric information in person identification can be separate into two types; the person recognition, and personal identification. Both of them are quite similar except the objectives are different. For the person recognition, the main objective is that learning machine systems can distinct persons that machines have learned that person's information before. In this case, machines have to learn large database that contain many people information and can recognize each person from others. Besides the person identification is the learning machine that have learned only one person information to respond only to that person. For the second case, there are many products that response to that kind of problem nowadays. Besides, many researchers give interests in the recognition system more than another. There are quit many researches about the biometric information. The most popular is the fingerprint recognition. From the ancient time, in China, they also use the fingerprint to identify people by the official. Until now the curiosity of the human increases parallel with the advancing technology. From using only fingerprint, the other parts of human body can also use as biometric information such as face, palmprint, and iris. Moreover teeth also use as information in recognition system. There are many works that have experiment on teeth recognition. These ideas inspired the creation of novel biometric information. As people known, there are many cosmetic surgeries. Other biometric are human tissues which can be changed by many causes. Besides teeth are bones which will not reshaped in long time. This is the reason in doing research about teeth recognition system. In addition there are other recognition systems that have the interesting algorithm. Those research works can be use as the references in starting the research about the recognition. Although the biometric information is different in human term, for the learning machine, they are just some kind of data. The learning machine will learn the feature part until they can distinct each data from others. Later they can recognize the same pattern of data but from different objects.

2.1. Using several frame images

The teeth recognition system is researched widespread however their methodology and principle are quit difference. Don-Ju Kim et al. proposed Teeth recognition based on multiple attempts in mobile device which is published in journal of Network and Computer applications. Their work is a novel recognition algorithm that using teeth-image based on multiple attempts. Usually most of the image-based biometrics systems utilize a single image during the recognition process. However, these algorithms sometimes fail to recognize users in practical situation due to false-detected or undetected object. Their hypothesis is that the use of several frame images, rather than just one, could improve the accuracy of recognition processes. Their algorithm is quite good. However, their training and testing database is not large. Only three and five images per person are collected. The total number of person in each experiment group is five. The selection process is based on differential image entropy (DIE). This will compute the differential between current frame image and an averaged teeth image. Because they use multi frame image they have to select the subjects that will be fed on to the recognition system. They compared the embedded hidden Markov model as the recognition with the principal component analysis and linear discriminant analysis.

According to these researchers, they also proposed another work that can develop on the mobile device. They considered that smart-phones are vulnerable to theft and loss due to their small size. A simple and convenient authentication system is required to protect private information stored in the mobile device. They proposed a multimodal biometric authentication approach using teeth image and voice as biometric traits. The individual matching scores obtained from the teeth image and voice are combined using a weighted-summation operation, and the fused-score is utilized to classify an unknown user into the acceptance or rejection. The teeth recognition part is the same as previous work except the voice recognition which is added to this work. They confirmed that the performance of the proposed system is better that the performance obtained using teeth only. However using voice is also include noise from

around the device. This may cause failure in some situations. If we cannot authenticate to our own mobile device, sometimes it cause problems.

2.2. Using single image

The multiple attempts teeth recognition gives high performance but the system has to select the images before fed to the classification step. This may take long time in the processing and complex in implementation. Besides there are many other recognition system that use single image based in the procedures. The one that use the well-known algorithm in their method is proposed by Supaporn Kiattisin and Chokemongkon Nadee, published Improved PCA and LDA-Based personal Identification Method at the third international symposium on biomedical engineering 2008. They used the Principal Component Analysis (PCA) system to identify person by the input image. In addition, their improved method was handled with the reflection and orientation by adding more step before fed the image to the classification step.

From reviewing, there are many other researches that does not mention above about using the Principal Component Analysis system in both feature extraction and classification. In the face recognition, they also use the Principal Component Analysis as classification. However there is a technique that is similar to the Principal Component Analysis in extracting feature and reducing the dimension of the image to remain only feature vector. That technique will be mentioned in the next chapter. There are other features that satisfied to be added in feature vector to advance the classification performance.

CHAPTER III

THEORITICAL BACKGROUND

Teeth-image recognition has the procedure as other biometric information recognition. To recognize object from the image usually uses the computer vision techniques. Basically a computer vision system processes images acquired from an electronic camera, which is like the human vision system where the brain processes images derived from the eyes. There are now many vision systems in routine industrial use: cameras inspect mechanical parts to check size, food is inspected for quality, and images used in astronomy benefit from computer vision techniques. In this studies, biometrics using computer vision include teeth recognition by the 'pattern' of their teeth.

This research involved many techniques to do teeth-image recognition. The procedure is separate to three segments. Begin with image preparation, before doing recognition each image have to convert to a computer understandable format. The second part is feature extraction. To extract important feature from the prepared image this procedure is involved. The last part is the classifying by classification models. Four basic classification models were compared their performances.

3.1. Image Preparation

3.1.1. Gray scale image

To make computer recognize the feature object from image can be done by the same inherent property of the human eye, known as Mach band affects the way we perceive images. These are illustrated in Figure 3.1 and are the darker bands that appear to be where two stripes of constant shade join. By assigning values to the image brightness levels, the cross-section of plotted brightness is shown in Figure 3.1(a). This shows that the picture is formed from stripes of constant brightness. Human vision perceives an image for which the cross-section is as plotted in Figure 3.1(c). These Mach bands do not really exist, but are introduced by your eye. The bands arise from overshoot in the eyes' response at boundaries of regions of different intensity (this aids

us to differentiate between objects in our field of view). The real cross-section is illustrated in Figure 3.1(b). Note also that a human eye can distinguish only relatively few grey levels. It actually has a capability to discriminate between 32 levels (equivalent to five bits) whereas the image of Figure 3.1(a) could have many more brightness levels. This is why your perception finds it more difficult to discriminate between the low intensity bands on the left of Figure 3.1(a).

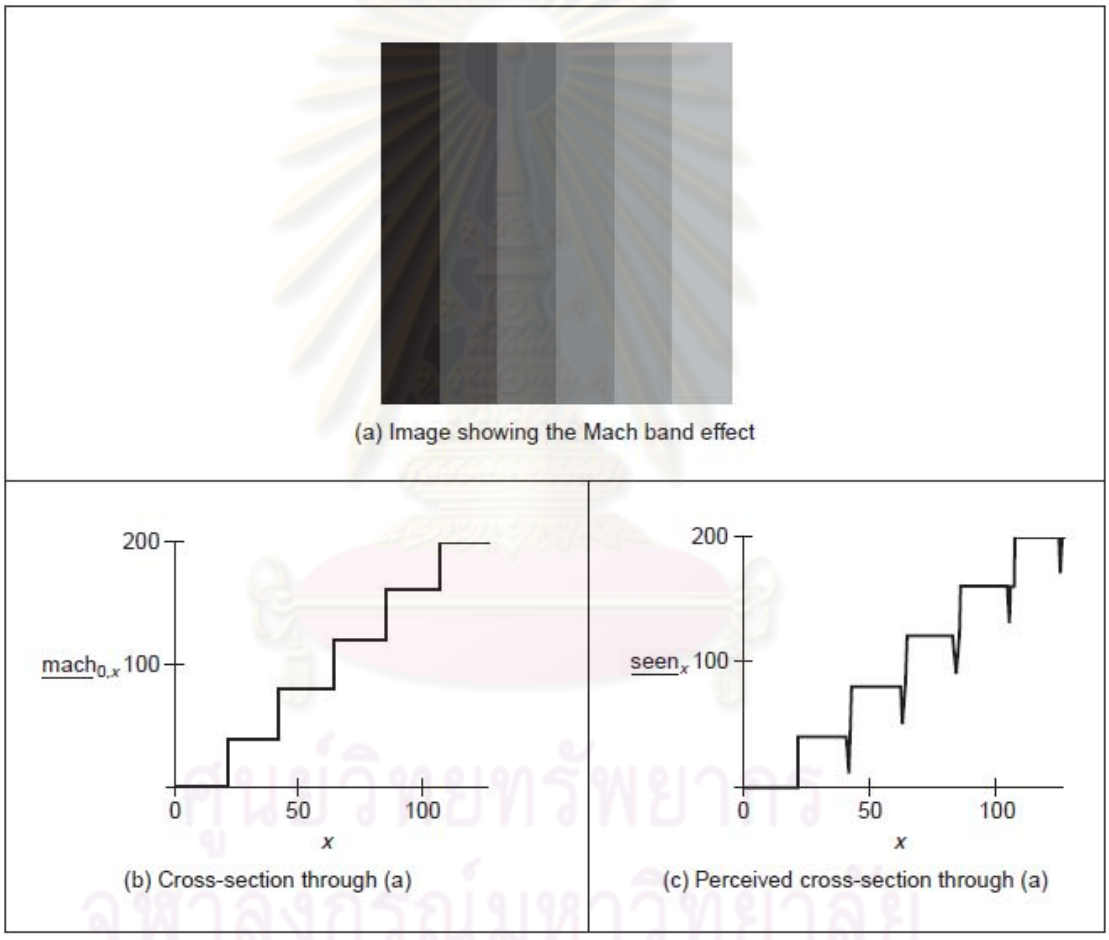


Figure 3.1 Illustrating the Mach band effect

According to previous paragraph, before feature extraction procedure is processed, the image preparation is require to be done first. Same as above theory, each image in this research is converted to gray scale image first. Although the computer can see the object in the image as human being, they might take long time to do classifying task in this domain. To perform the task faster, the domain was changed

to another domain, frequency domain. Among the frequency domain, there are few transformations in considering. The one that is used in this dissertation is shown below.

3.1.2. Discrete Cosine Transform(DCT)

The discrete cosine transform (DCT) represents an image as a sum of sinusoids of varying magnitudes and frequencies. The DCT has the property that, for a typical image, most of the visually significant information about the image is concentrated in just a few coefficients of the DCT. For this reason, the DCT is often used in image compression applications. For example, the DCT is at the heart of the international standard lossy image compression algorithm known as JPEG. In this research, the discrete cosine transform is considered for computer performing faster by converting image to frequency domain.

The two-dimensional DCT of an M-by-N matrix A is defined as follows.

$$F_{ij} = \alpha_i \alpha_j \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\left[\frac{\pi(2m+1)i}{2M}\right] \cos\left[\frac{\pi(2n+1)j}{2N}\right], \quad 0 \le p \le M-1, 0 \le q \le N-1, \tag{1}$$

Where $\alpha_i = \begin{cases} \frac{1}{\sqrt{M}} & , i = 0 \\ \sqrt{\frac{2}{M}} & , 1 \le i \le M-1 \end{cases}$, $\alpha_j = \begin{cases} \frac{1}{\sqrt{N}} & , j = 0 \\ \sqrt{\frac{2}{N}} & , 1 \le j \le N-1 \end{cases}$

The values F_{ij} are called the DCT coefficients of A. (Note that matrix indices in MATLAB always start at 1 rather than 0; therefore, the MATLAB matrix elements $A(1,1)$ and F (1,1) correspond to the mathematical quantities A_{00} and F_{00}, respectively.)

The image preparation is complete at this step. Before the objects are fed to classification models for classifying step, the object which is frequency domain have to be analyzed and extracted the features. The most important in this research is feature extraction where two features is involved. The proposed feature, called hybrid features,

separates into global feature and local feature. In this research, the global feature is considered from some existing techniques that are discussed next.

3.2. Feature extraction

There are many techniques that are considered as feature extraction techniques. Principal Component Analysis (PCA) is often used for feature extraction. This technique is considered as candidate with another technique that is used in this research, Singular Value Decomposition (SVD). One goal of these techniques is to reduce dimension of the object. Although the discrete cosine transform had been used to convert domain to frequency for performing faster, the dimension of the object is yet high. Problems arise when performing recognition in a high-dimensional space. To prevent the curse of dimensionality occurs in classifying, Principal Component Analysis and Singular value Decomposition are applied to be feature extraction. Both techniques are discussed below.

3.2.1. Principal Component Analysis (PCA)

Principal Component Analysis is a technique to identify the patterns in data and express the data in such a way as to emphasize their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data.

For PCA to work properly, each of the data dimensions have to be subtracted the mean. The mean subtracted is the average across each dimension. So, all the x values have $\bar{x}$ (the mean of the x values of all the data points) subtracted, and all the y values have $\bar{y}$ subtracted from them. This produces a data set whose mean is zero. Table 3.1 shows the sample data that has x values and y values. Table 3.2 shows the adjusted data which are subtracted with the mean of each value. The subtracted mean values are shown below.

Table 3.1 Sample data

x values	y values
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

Table 3.2 Adjusted data

x values	y values
0.69	0.49
-1.31	-1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

$$\bar{x} = \frac{2.5+0.5+2.2+1.9+3.1+2.3+2+1+1.5+1.1}{10} = 1.81.$$

$$\bar{y} = \frac{2.4+0.7+2.9+2.2+3+2.7+1.6+1.1+1.6+0.9}{10} = 1.91.$$

The adjusted data is used to calculate the covariance matrix. This is done in the same way of variance. Basically variance is used to calculate 1-dimensional data.

$$\mathrm{var}(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}. \qquad (2)$$

For more than one dimension, the covariance is used instead. Since both of them are the same technique, their formulas are similar. The equation below is for calculating covariance for 2-dimension.

$$C = \begin{pmatrix} \text{cov}(x,x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{cov}(y,y) \end{pmatrix}. \quad (3)$$

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}. \quad (4)$$

Where n equals to 10, in this sample case. To calculate $\text{cov}(x,x)$ is the same as $\text{var}(x)$ since there is calculation in one dimension. According to the equation (4), we knew that $\text{cov}(x,y)$ equals to $\text{cov}(y,x)$. Table 3.3 below shows the result in covariance of the sample data.

Table 3.3. 2-dimensional data set and covariance calculation

i	x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2.5	2.4	0.69	0.49	0.4761	0.2401	0.3381
2	0.5	0.7	-1.31	-1.21	1.7161	1.4641	1.5851
3	2.2	2.9	0.39	0.99	0.1521	0.9801	0.3861
4	1.9	2.2	0.09	0.29	0.0081	0.0841	0.0261
5	3.1	3	1.29	1.09	1.6641	1.1881	1.4061
6	2.3	2.7	0.49	0.79	0.2401	0.6241	0.3871
7	2	1.6	0.19	-0.31	0.0361	0.0961	-0.0589
8	1	1.1	-0.81	-0.81	0.6561	0.6561	0.6561
9	1.5	1.6	-0.31	-0.31	0.0961	0.0961	0.0961
10	1.1	0.9	-0.71	-1.01	0.5041	1.0201	0.7171
Average					0.616556	0.716556	0.615444

From the table 3.3, the covariance matrix of this sample data is

$$\text{cov} = \begin{pmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{pmatrix}.$$

Since the covariance matrix is square, the eigenvectors and eigenvalues can be calculated for this matrix. These are rather important, as they represent the useful information about the data. Basically to find the eigenvalues for low dimension $n \times n$ matrix, in this case, n is equal to 2, the eigenvalues and eigenvectors of the covariance matrix is shown below.

$$eigenvalues = \begin{pmatrix} 0.0491 \\ 1.284 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -0.7352 & 0.6779 \\ 0.6779 & 0.7352 \end{pmatrix}$$

After obtaining the eigenvectors, the highest eigenvaluse is choosen then the eigenvector of that eigenvalue multiplies with the sample data. In fact, it turns out that the eigenvector with the highest eigenvalue is the principle component of the data set. It is the most significant relationship between the data dimensions. Finally the feature vector of sample data is reduced to one-dimension. This feature vector is fed to classification model to do classification.

3.2.2. Singular Value Decomposition (SVD)

The singular value decomposition is the technique in the linear algebra. This is very useful technique in data analysis and visualization. It emphasizes initial characterization of the data and represents the data using a smaller number of variables. In this work, the singular value decomposition is proposed as feature extraction technique along with color histogram and teeth-width ratio.

The singular value decomposition of a matrix factors an mxn matrix A into the form in equation below.

$$A = USV^T. \tag{5}$$

Where U is an mxm orthogonal matrix; V an nxn orthogonal matrix, and S an mxn matrix containing the singular values of A along its diagonal.

For a simple example, suppose A matrix is $4x2$ dimensions contain values as below.

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Matrix V from (5) contains the corresponding eigenvectors of A^TA. After obtaining eigenvalues, they are rearranged the values in descending order. The matrix V is the eigenvector that rearranged the column following the order of eigenvalues.

$$eigenvalues \text{ of } A^TA = \begin{bmatrix} 0.1339 & 0 \\ 0 & 29.8661 \end{bmatrix}$$

$$, eigenvectors \text{ of } A^TA = \begin{bmatrix} -0.9145 & -0.4046 \\ 0.4046 & -0.9145 \end{bmatrix}$$

After rearranging the values, the set of eigenvectors associated below as V matrix.

$$V = \begin{bmatrix} -0.4046 & -0.9145 \\ -0.9145 & 0.4046 \end{bmatrix}$$

Next, the singular value of A matrix is defined from the square root of the corresponding eigenvalues of the matrix A^TA in the same order as V. Earlier it was defined to be diagonal matrix which is mxn dimensions. Let S_1 be a square rxr matrix containing the singular values of A along its diagonal. Therefore the matrix S is placed with the zero singular values at the end of the diagonal matrix, S_1. The correct

dimension is padded with $m-r$ rows of zeros and $n-r$ columns of zeros. In this sample, the S_1 matrix is padded with 2 rows and 0 columns as below.

$$S = \begin{bmatrix} 5.465 & 0 \\ 0 & 0.3659 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

For last matrix U, this matrix is obtained from eigenvectors of AA^T with the method as obtaining matrix V. The eigenvalues and eigenvectors of AA^T is illustrated below.

$$eigenvalues \text{ of } AA^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1339 & 0 \\ 0 & 0 & 0 & 29.8661 \end{bmatrix},$$

$$eigenvectors \text{ of } AA^T = \begin{bmatrix} 0 & 0 & -0.576 & -0.8174 \\ 0 & 0 & 0.8174 & -0.576 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So the rearrange of the eigenvectors of AA^T in the same order of as their respective eigenvalues is matrix U.

$$U = \begin{bmatrix} -0.8174 & -0.576 & 0 & 0 \\ -0.576 & 0.8174 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The proposed method in this work is obtained only feature vector which is from the main diagonal of singular values matrix S. This technique helps compressing the image to only important pattern for reducing time in classifying.

3.2.3. Gray scale and color Histogram

Histogram technique is usually use in image processing and computer vision. Histograms are used to plot density of image. For the gray scale histogram, the density is calculated from the density of gray image which has range between 0 and 255. The histogram obtained by counting the density of each pixel. Consequently the gray scale histogram feature vector has length 256.

For the color histogram, a pixel is separated into three colors which are red, green, and blue. This feature vector was separately counted the density then the feature vector is 768 long.

3.3. Classification models

The classification models play an important role in this research. More over the classification models are used in many works that involve the pattern recognition both in statistic and computer vision. Basically the classification model analyzed the characteristic of the data-set or found the object in the image by recognizing the pattern.

In the experiment, four existing classification models consisting of Naïve Bayes model, K-nearest neighbor classifier, multilayer perceptron with backpropagation learning and Levenberg Marquardt training algorithm and Resilient backpropagation classifiers were under the consideration.

3.3.1. Naïve Bayes classifier

Naïve Bayes classifier is a probabilistic classifier based on independent feature model. Meaning the Naïve Bayes classifier is designed for use when features are independent of one another within each class, but it appears to work well in practice even when that independence assumption is not valid. It classifies data in two steps:

1. Training step: Using the training samples, the method estimates the parameters of a probability distribution, assuming features are conditionally independent given the class.
2. Prediction step: For any unseen test sample, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test sample according the largest posterior probability.

The class-conditional independence assumption greatly simplifies the training step since you can estimate the one-dimensional class-conditional density for each feature individually. While the class-conditional independence between features is not true in general, research shows that this optimistic assumption works well in practice. This assumption of class independence allows the Naïve Bayes classifier to better estimate the parameters required for accurate classification while using less training data than many other classifiers. This makes it particularly effective for datasets containing many features.

Naïve Bayes classification is based on estimating $P(X \mid Y)$, the probability or probability density of features X given class Y. Before dataset are fed to the Naïve Bayes classification model, the dataset must be provided the distributions to estimate the parameters for a feature's distribution. In this research, the normal distribution is considered. This distribution is appropriate for features that have normal distributions in each class. The Naïve Bayes classifier estimates a separate normal distribution for each class by computing the mean and standard deviation of the training data in that class.

The classifier is based on the Bayes' Theorem or Bayesian inference. This method uses evidence to calculate the probability that the hypothesis may be true. The equation below shows the simple formula that describes the Bayes' Theorem.

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}. \quad (6)$$

Where H : The hypothesis.

E : The evidence that has been observed.

$P(H)$: The prior probability of H that was inferred before new evidence is available.

$P(E \mid H)$: The condition probability of seeing the evidence E if the hypothesis H is true.

$P(E)$: The priori probability of witnessing the new evidence E under all possible hypotheses. It can be calculated as the sum of the product of all probabilities of any complete set of mutually exclusive hypotheses and corresponding conditional probabilities:

$$P(E) = \sum_{i=1}^{n} P(E \mid H_i)P(H_i).$$

$P(H \mid E)$: The posterior probability of H given E and the new estimate of the probability that the hypothesis H is true, taking the evidence E into account.

To compare know and predicted classes of observations, the confusion matrix is used to tabulate misclassifications. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in a known class. The diagonal is the results that correctly classified. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes. Table 3.4 shows an example of confusion matrix.

Table 3.4. An example of confusion matrix

		Predicted		
		Class 1	Class 2	Class 3
Known	Class 1	5	3	0
	Class 2	2	3	1
	Class 3	0	2	11

3.3.2. k-Nearest neighbor

The k-nearest neighbor algorithm is a classifying method based on closest training examples in the feature space, in pattern recognition. The k-nearest neighbor is called lazy learning where the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbor algorithm is the simplest among all of machine learning algorithm. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors, usually k is small positive integer. In this experiment, k is set to 1 then the object is simply assigned to the class of its nearest neighbor. A Euclidean distance measure is used to calculate how close each member of the training set is to the object that is being examined.

For simple example, suppose sample data and training data are set as below with assigned classes.

Table 3.5. Sample data with class assigned

Class	x	y
1	0.9	0.8
2	0.1	0.3
3	0.2	0.6

Table 3.6. Training data

x	y
0	0
0.5	0.5
1	1

The Euclidean distance measure is described in the equation (7). After each data in the training data is calculated the distance between each data each class, the lowest distant between training and sample data means the nearest.

$$dist(s,t) = \sqrt{(s_x - t_x)^2 + (s_y - t_y)^2}. \quad (7)$$

Where S_x : A sample data in x column.

S_y : A sample data in y column.

t_x : A training data in x column.

t_y : A training data in y column.

Table 3.7. The table show the Euclidean distance between each class member and training data (0,0)

Class	x	y	Euclidean distant
1	0.9	0.8	$\sqrt{(0.9-0)^2 + (0.8-0)^2} = 1.2042$
2	0.1	0.3	$\sqrt{(0.1-0)^2 + (0.3-0)^2} = 0.3162$
3	0.2	0.6	$\sqrt{(0.2-0)^2 + (0.6-0)^2} = 0.6325$

From the table 3.7, the training data (0,0) is closest to sample data of class 2, so that training data belong to class 2. Because k is set to 1, the training data can selected only one nearest neighbor. Besides the large data set, k may be set larger to reduce misclassified. In this simple example, each class has only one sample data so training data can select only one closest. For large data set, suppose k is set to 4, the

first 4 closest neighbors are selected. The training data class will be chosen from the class that has greatest number of closest neighbors among those 4 neighbors. The Euclidean distant measure for the rest of data is shown below. The result is shown in the table 3.10.

Table 3.8. The table show the Euclidean distance between each class member and training data (0.5,0.5)

Class 3.3.3.	x	y	Euclidean distant
3.3.1.4.	0.9	0.8	$\sqrt{(0.9-0.5)^2+(0.8-0.5)^2}=0.5$
2	0.1	0.3	$\sqrt{(0.1-0.5)^2+(0.3-0.5)^2}=0.4472$
3	0.2	0.6	$\sqrt{(0.2-0.5)^2+(0.6-0.5)^2}=0.3162$

Table 3.9. The table show the Euclidean distance between each class member and training data (1,1)

Class	x	y	Euclidean distant
1	0.9	0.8	$\sqrt{(0.9-1)^2+(0.8-1)^2}=0.2236$
2	0.1	0.3	$\sqrt{(0.1-1)^2+(0.3-1)^2}=1.1402$
3	0.2	0.6	$\sqrt{(0.2-1)^2+(0.6-1)^2}=0.8944$

Table 3.10. Result table of the k-Nearest Neighbor classification

x	y	Predicted Class
0	0	2
0.5	0.5	3
1	1	1

3.3.3. Resilient backpropagation training algorithm

The Resilient backpropagation algorithm is the training algorithm in multilayer perceptron. The backpropagation is the generalization of the Widrow-Hoff learning rule to multilayer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined. Networks with biases, a linear transfer function as hidden layer and output layer are capable of approximating any function with continuities. Figure 3.2 shows the linear transfer function in graph.

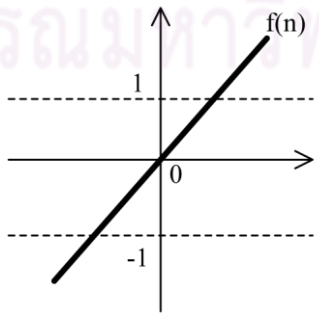


Figure 3.2. Linear transfer function

Standard backpropagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks. There are a number of variations on the basic algorithm that are based on other standard optimization techniques, such as conjugate gradient and Newton methods.

Basically feedforward networks often have one or more hidden layers of neurons followed by an output layer of linear neurons. Multilayers of neurons with linear transfer functions allow the network to learn linear relationships between input and output vectors. The linear output layer let the network produce values outside the range -1 to +1. The figure 3.3 illustrates the neural network architecture which is used in this research that has two linear hidden layers and a linear output layer.

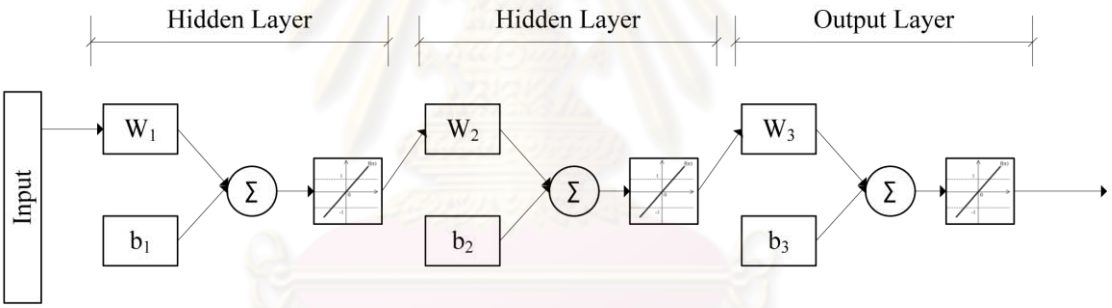


Figure 3.3. Neural network architecture

For the time being two-layer feed-forward networks, one hidden layer and an output layer, can potentially learn virtually any input-output relationship, feedforward networks with more layers might learn complex relationships more quickly. As complex dataset in this research, the proposed neural network model has two hidden layers and an output layer to deal with the complexity of the dataset. There are generally two steps in this algorithm; training and test step. First the neural network has to learn on the examples of desired behavior. The desired behavior is summarized by a set of input and output. The neural network calculates desires change to their weights ($\mathbf{W}$) and biases ($\mathbf{b}$), given input vectors ($\mathbf{p}$). The desired output vector contains values of either -1 or 1, because the neural network has the linear transfer function.

Once the neural network weights and biases are initialized, the neural network is ready for training. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function. The default performance function for feedforward networks is mean square error; the average squared error between the neural network outputs and the desired outputs. This algorithm uses the gradient of the performance function to determine how to adjust the weights to minimize performance. The gradient is determined using a technique called backpropagation, which involves performing computations backward through the network. Basically this training algorithm, the weights are moved in the direction of the negative gradient. Therefore there is a problem when using steepest descent in training a multilayer network, because the gradient can have a small magnitude and cause small changes in the weights and biases, even though the weights and biases are far from their optimal values.

The Resilient backpropagation (Rprop) training algorithm is to eliminate these harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative can determine the direction of the weight update. That means the magnitude of the derivative has no effect on the weight update. The size of the weight change is determined by a separate update value. The update value for each weight and bias is increased whenever the derivative of the performance function with respect to that weight has the same sign for the two successive iterations. The update value is decreased whenever the derivative with respect to that weight changes sign from the previous iteration. If the derivative is zero, the update value remains the same. Whenever the weights are oscillating, the weight change is reduced. If the weight continues to change in the same direction for several iterations, the magnitude of the weight change increases.

The advantage of the Resilient backpropagation (Rprop) training algorithm is generally much faster than the standard steepest descent algorithm. Besides Resilient backpropagation (Rprop) training algorithm also has the nice property that it requires only a modest increase in memory requirements.

3.3.4. Multilayer perceptron with Levenberg-Marquardt learning

The Levenberg-Marquardt algorithm is the learning algorithm in the neural network backpropagation. This algorithm also increases the speed in learning as the Resilient backpropagation. The Levenberg-Marquardt use the quasi-Newton methods to perform the training faster.

Originally Newton's method is an alternative to the conjugate gradient methods for fast optimization. It uses the Hessian matrix which is the second-order derivatives of the performance function of the weights and biases. Newton's method often converges faster than conjugate gradient method. Unfortunately, it is complex to compute the Hessian matrix for feedforward neural networks. There is an algorithm that is based on Newton's method, but which doesn't have to calculate second derivatives. These are called quasi-Newton or secant methods. The neural networks update an approximate Hessian matrix at each iteration of the algorithm. The update is computed as a function of the gradient. Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks) then the Hessian matrix can be approximated as

$$H = J^T J. \tag{8}$$

and the gradient can be computed as

$$g = J^T e. \tag{9}$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of the network errors. The Jacobian matrix can be computed through standard backpropagation techniques that are much less complex than computing the Hessian matrix.

The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton's method update

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e. \quad (10)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm.

This algorithm appears to be the fastest method for training moderate-sized feedforward neural networks (up to several hundred weights). The best type of problem for the Levenberg-Marquardt algorithm is a function approximation problem where the network has fewer than hundred weights and the approximation must be very accurate. In general, the Levenberg-Marquardt algorithm does not perform as well on pattern recognition problems as it does on function approximation problems. The Levenberg-Marquardt algorithm is designed for least squares problems that are approximately linear. In this research, the Levenberg-Marquardt algorithm is proposed as the classification model in the system notwithstanding. Thus the transfer function in the hidden layer is linear transfer function.

3.4. Cross-validation

The cross validation is used to generalize an independent data set. This technique uses when setting the data set to estimate how accurately a predictive model will perform in practice. The data set is apart into equally subset. The training set is used to perform the analysis. The validation set is used to validate the analysis.

This work is involved the k-fold cross validation. The original sample is randomly partitioned into *K* subsamples. Of the *K* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ subsamples are used as training data. The cross-validation process is then repeated *k* times (the *folds*), with each of the *K* subsamples used exactly once as the validation data. The *k* results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. 5-fold cross-validation is considered in this work.

The detail of the proposed system is described in the next chapter. From the preprocessing until the classifying, they are described step by step with flow diagram of this system.

CHAPTER IV

PROPOSED METHOD

In this part, our proposed features and model are described. The overall system is depicted in flow diagram as shown in Figure 4.1. It can be separated into two consecutive parts: the proposed features and classification model. From the various models evaluation, using singular value decomposition and color histogram as input features yields good results. Besides the database is larger, the accuracy becomes lower. Thus other feature is considered to be added. In the second part, the local feature with the properties of compact size and distinguishably is considered. Such feature is teeth width ratio. For every single smiling image, that teeth is visible, the most visible part is the upper front teeth that wide spread along with person lips. Therefore the width of the teeth is collected. In the other hands, the height of the teeth cannot be considered due to its invisibility in some cases. After hybrid features merged from global and local features are extracted, they are fed into the classification model. The classification model is considered from several models. The one that is simple like neural network backpropagation is selected. Among the neural network backpropagation, there are many learning algorithm to be considered. The fast algorithm is considered first. Among the fast learning algorithm, the evaluation experiment is analyzed. The one selected to be proposed is the Levenberg-Marguardt learning algorithm, based on highest accuracy.

In this chapter, the explanation of proposed method is written. First, the feature extraction is explained. The hybrid features is come from global feature and local feature. Each feature is described with the method and the calculation. Finally, the classification model, neural network backpropagation with Levenberg-Marquardt learning algorithm is described along with its architecture model, which is used in this experiment.

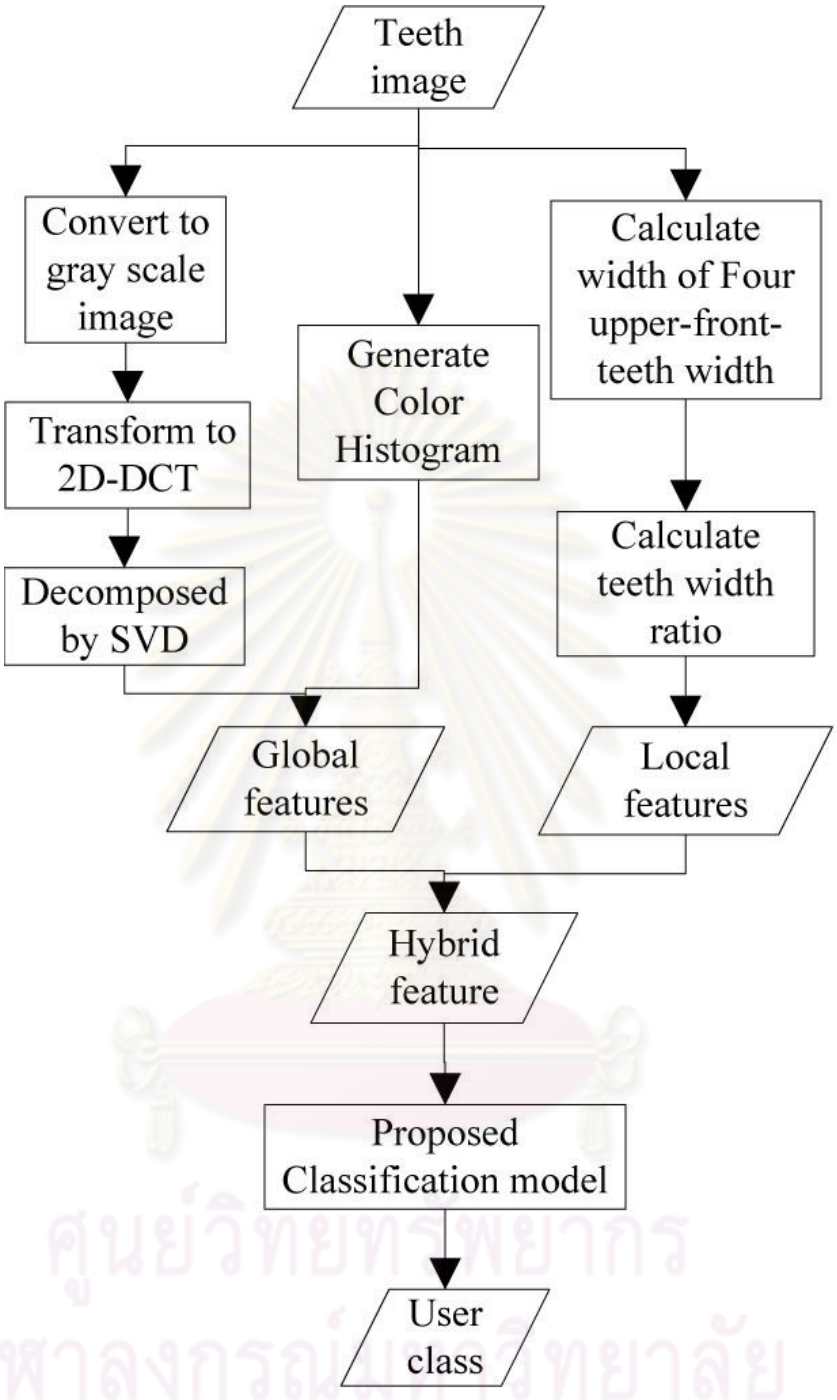


Figure 4.1. Flow Diagram

4.1.Feature Extraction

The extraction process can be divided into two major parts corresponding to global features and local features.

4.1.1. Global Feature

For this part, two-Dimensional Discrete Cosine Transform (2D-DCT) is proposed to extract DCT coefficients of teeth image. The 2D-DCT is selected because it has been widely used for extracting a unique set of features for each image in a training set. Due to a property of linear separable transform, 2D-DCT can be processed in term of a traditional DCT (1D-DCT) by performing along a single dimension followed by a one-dimensional DCT in another dimension. For an input image *A* of size ($M \times N$) and the transformed image *B*, the 2D-DCT can be performed by (1), where *M* and *N* are number of rows and columns of image *A*, respectively.

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos[\frac{\pi(2m+1)p}{2M}] \cos[\frac{\pi(2n+1)q}{2N}], \quad (11)$$
$$0 \le p \le M-1, 0 \le q \le N-1,$$

Where $$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}} & , p = 0 \\ \sqrt{\frac{2}{M}} & , 1 \le p \le M-1 \end{cases} , \alpha_q = \begin{cases} \frac{1}{\sqrt{N}} & , q = 0 \\ \sqrt{\frac{2}{N}} & , 1 \le q \le N-1 \end{cases}$$

The image of transform coefficients can be shown as Figure 4.2.

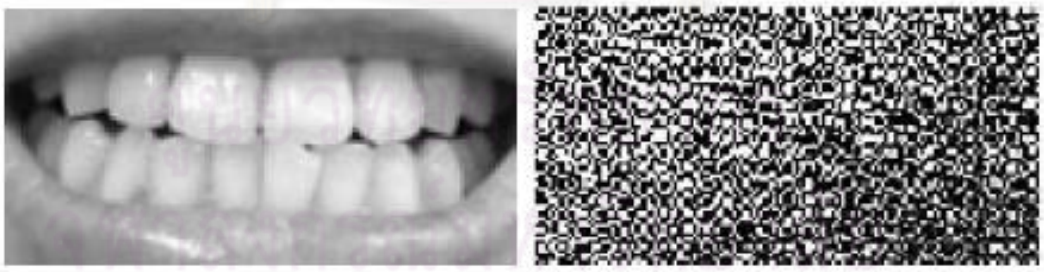


Figure 4.2. Original grey scale image (left) and its 2D-DCT (right)

From the image of 2D transform coefficients, Singular Value Decomposition (SVD) is applied in the next step. Only diagonal entries in the decomposition matrix are used as the first set of features for our system. This aims to

extract the unique features in compact size for training process with less time consumption. To compute SVD, the DCT coefficient matrix can be formed by

$$A = USV'. \tag{12}$$

where A : 2D-DCT coefficient matrix of size m x n

U : orthogonal columns matrix of size m x r

S : orthogonal columns matrix of size r x r

V : orthogonal columns matrix of size n x r

The composition of 2D-DCT coefficient matrix can be explained in Figure 4.3.

$$\begin{bmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mr} \end{bmatrix} \begin{bmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{1r} & & v_{rn} \end{bmatrix}$$

Figure 4.3. Singular Value Decomposition matrix

The second set of global features is obtained from color image histogram. Teeth image are separated into three color channels: red, green, and blue. Then, for each color channel, the frequency of each intensity value is accounted to perform a histogram vector. Combining three histogram vectors with concatenation yields a vector of color histogram. Finally, a set of diagonal entries in singular value matrix and a color histogram will be merged together for creating a complete set of global features

4.1.2. Local Feature

Local features are the features that can be obtained from some specific locations in the image. For this part, the width of each of four upper front teeth is considered. The reason of selecting a group of upper front teeth is that most of image

captured when people smile contains these teeth. To extract local features, suppose that there is a line across four upper front teeth. The intensity values along this line can be projected as shown in Figure 4.4. and 4.5. It can be obvious that edge of two adjacent teeth performs local minimum in intensity projection.

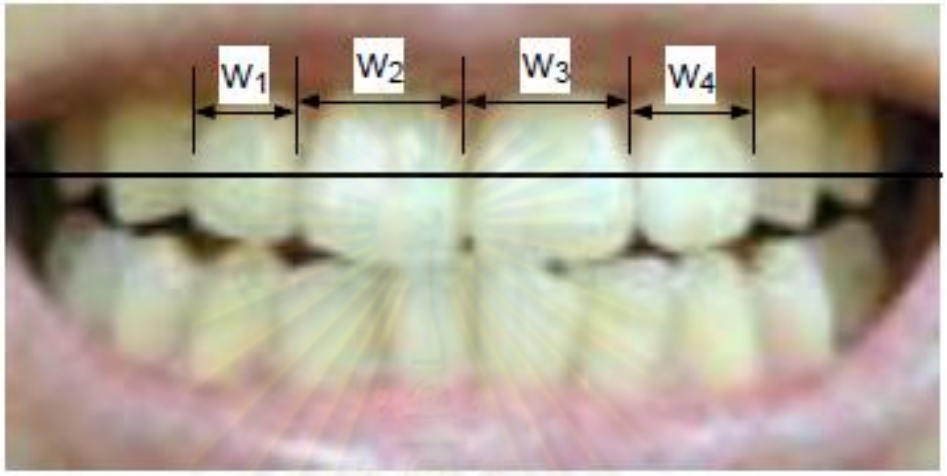


Figure 4.4. The width of each of four upper front teeth

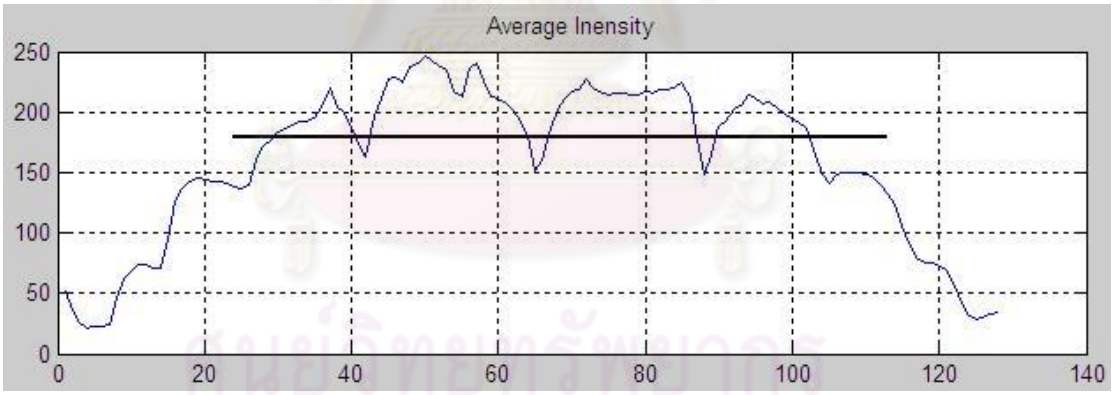


Figure 4.5. The intensity of the front teeth

Then, the width of each of four upper front teeth, w_i , is extracted from measuring the distance between two points causing local minimum. After that, the normalized width, $\hat{W}_i$, of each teeth can be calculated from the equation.

$$\hat{W}_i = \frac{w_i}{\sum_{i=1}^{4} w_i} \tag{13}$$

$$\hat{W}_i \text{ for } i = 1, 2, 3, 4.$$

After extraction of global and local features, these features are combined into a vector of hybrid features and used as input of classification model. Figure 4.6 shows the proposed vector.

Figure 4.6. Feature vector

4.2. Classification Model

Due to the purpose of reducing time consumption in training step and during the process, multilayer perceptrons backpropagation with Levenberg-Marquardt training algorithm is selected as the simple model for our system. For the multilayer perceptrons, the output at unit i in layer k + 1 is:

$$n^{k+1}(i) = \sum_{j=1}^{k} w^{k+1}(i, j) a^{k}(j) + b^{k+1}(i). \tag{14}$$

where *w(i,j)* is the weight connecting from unit *j* to unit *i*, *a* is the output from layer *k*, and *b* is the bias at layer *k*+1 of neural network.

The output at unit *i* in layer *k*+1 is:

$$a^{k+1}(i) = f^{k+1}(n^{k+1}(i)). \tag{15}$$

Then,

$$a^{k+1}(i) = f^{k+1}(\sum_{j=1}^{k} w^{k+1}(i,j)a^{k}(j) + b^{k+1}(i)). \quad (16)$$

To find the recurrence relation at the final layer, the first derivatives of the network errors and updating function for each layer must be found.

For the Levenberg-Marquardt algorithm, and approximation of the Hessian matrix, the matrix of second derivatives of the updating function which difficult to compute by ordinary analytical methods, so this matrix can be defined using the Jacobian matrix as (7).

$$H = J^{T}J. \quad (17)$$

The Jacobian matrix containing first derivatives of the network errors with respect to parameter vector x in each layer is used to minimize the updating function. The parameters can be updated by:

$$x_{k+1} = x_k - [J^{T}J + \mu I]^{-1} J^{T} e. \quad (18)$$

Where e is a network error vector, the μ is a constant to control the speed of parameter learning. In this experiment, we set parameter μ to 0.001 with increasing adaptive value as 5 and decreasing adaptive value is 0.05. There are 2 hidden layers with 3 and 4 neurons respectively in each layer. Each node is linear transfer function. For output layer, the linear transfer function also considered along with target is set to be between -1 and 1. This model architecture is shown Figure 4.7.

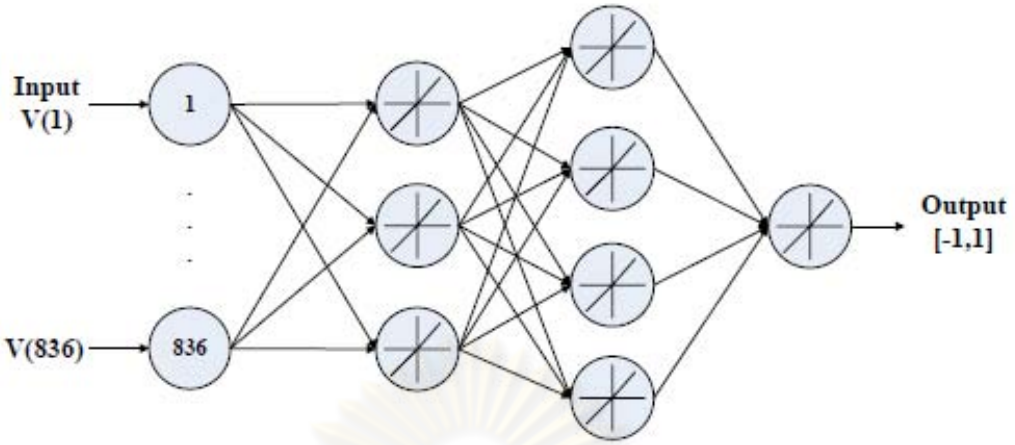


Figure 4.7. Architecture of Neural network backpropagation model used in this experiment

CHAPTER V

EXPERIMENT

In the experiment, the images that have been used in the dataset are taken by digital camera, Nikon D3000 with 18-55 VR lens. The original image resolution is 1944 x 2896 pixels. From the collection, region of teeth are focused and manually located under the normalized size of 64 x 128 pixels. The example of the teeth images in the dataset is shown in Figure 5.1. There are 25 subjects in this experiment which are classified into 25 classes and 20 pictures are taken for each person. Therefore, the total number of teeth images is 500.

Figure 5.1. Example of an image in dataset

The experiment was separated into two parts. In the first part, various feature extractions are discussed with some of classification models to compare their performance. In this part, 10 persons with 20 pictures per each person were used to determine the performance. Totally, there are 10 classes for 10 persons. The all misclassified are false acceptance. Besides, four feature extraction techniques which were PCA, SVD, gray scale histogram, and color histogram were also considered as well.

5.1. First Experiment

In first part of experiment, the above feature extractions were fed to four classification models. The selected models are Naïve Bayes, k-Nearest Neighbor, Resillient propagation and the proposed model, multilayer perceptrons with Levenberg-Marquardt training.

In the experiment, the proposed feature is considered to evaluate with other features such as values from Principal Component Analysis. Principal Component Analysis can be used as classification model and feature extraction. In this case, to evaluate with the proposed feature, the Principal Component Analysis can be considered as feature extraction. Further from Principle Component Analysis and Singular Value Decomposition techniques, the color of the teeth is considered to be used to classify person. Thus the color histogram and gray histogram are considered to be compared with those two techniques. To ensure the result from the proposed method, only the color histogram is tested to compare with the proposed feature which including the Singular Value Decomposition and color histogram.

For classification models, the well-known Naïve Bayes classification is considered to this test. The k-nearest neighbor which is widely used in the people recognition systems is also included. Those models are selected to compare with the proposed neural network backpropagation models. Further to analyze the result in using neural network bakpropagation, two learning algorithm are selected to compare with those two models. Those learning algorithms are Resilient backpropagation and Levenberg-Marquardt algorithm.

For Naïve Bayes model, either principal components or singular values were independently selected. The ratio of training and testing was 75:25. Naïve Bayes is a probabilistic classifier with normal Gaussian distribution. Like other classifiers, the model attempts to classify teeth images into personal class. In the training process, all related parameters are estimated from the conditional probability of occurrence to form the function of probability distribution for independent features in the given class. In the

testing process, the model calculates the posterior probability of the teeth image for an appropriate class. This model yields high false acceptance rate because there is no reject class. In other words, all misclassified images are all images assigned into other class. The False Acceptance Rate (FAR) can be calculated by (9).

$$FAR = \frac{f_n}{f_n + f_p}. \tag{19}$$

where f_n : Number of False accepted Images

f_p : Number of True accepted Images

For k-Nearest Neighbor, this classifier classifies object based on closest samples in the feature space. In this work, k was equal to 1. In this classification, the unlabeled object is assigned the label which is most frequent among the k training samples nearest to the considered object. Five systems can be constructed based on this classifier. For each personal class, each image was set to be the labeled object.

The third model used to identify person to compare with our method was resilient backpropagation classifier. For this model, five feature extractions were all considered. Five-fold cross validation was also considered together with 75:25 ratios of training images and testing images. The momentum was initially set to 0.9. The learning rate and maximum epoch were 0.5 and 50, respectively. In addition the acceptance threshold for activation function was also 0.55. The R-prop differs to traditional backpropagation learning in term of weight tuning based on only the change of direction or sign of the derivative without the considering the magnitude of the derivative. The advantage of Resilient propagation is speed for training. The architecture of this model is the same as shown in Figure 4.7. Besides, only the algorithms for training are different.

For our algorithm, with the backpropagation learning and Levenberg-Marquardt training algorithm, dataset was separated for training and testing in the ratio of 75:25 with the inclusion of five-fold cross validation. The momentum was initially defined to 0.9. The parameter μ and maximum epoch were 0.001 and 50, respectively. The threshold of

acceptance for the activation function was set to be 55. In other words, the teeth image was accepted for the appropriate personal class if the output is greater than or equal to 0.55. Otherwise, the image was rejected from the class.

After images were classified with provided models, the accuracy rate and False Acceptance Rate (FAR) were calculated to comparison, as shown Table 5.1.

Table 5.1 : Result table of first part of experiment

Model	Accuracy rate (%)	FAR(%)
Levenberg Marquardt + SVD, color histogram (Proposed system)	**96**	**0**
Levenberg Marquard t+ PCA	76	11.63
Levenberg Marquardt + SVD	80	6.98
Levenberg Marquardt + SVD, gray histogram	90	6.25
Levenberg Marquardt + color histogram	94	4.08
Naïve Bayes + PCA	88	12
Naïve Bayes + SVD	84	16
Naïve Bayes + SVD, gray histogram	80	20
Naïve Bayes + SVD, color histogram	64	36
Naïve Bayes + color histogram	62	38
k-Nearest Neighbor + PCA	74.21	25.79
k-Nearest Neighbor + SVD	77.895	22.1
k-Nearest Neighbor + SVD, gray histogram	80	20
k-Nearest Neighbor + SVD, color histogram	83.684	16.316
k-Nearest Neighbor + color histogram	91.053	8.947
R-Prop + PCA	52	16.13
R-Prop + SVD	84	2.325
R-Prop + SVD, gray histogram	84	2.27
R-Prop + SVD, color histogram	60	0
R-Prop + color histogram	78	2.5

According from the comparison of results, the features based on SVD and color histogram and multilayer perceptron with backpropagation learning and Levenberg-

Marquardt algorithm were suitable for teeth recognition system. From the experiment, our proposed system gave the accuracy about 96% without any false acceptance. This means the SVD and color histogram feature can distinct each personal teeth-image. With the multilayer perceptron, backpropagation Levenberg-Marquardt learning algorithm, it can classify the features effectively without misclassifying wrong person. Due to the result, clearly, the features based on SVD and color histogram and multilayer perceotron with Levenberg-Marquardt learning was suitable to use in developing the teeth recognition system. Although the accuracy is about 96% without FAR, the performance might be improve after adding more specific features. The second part after this part is about the hybrid features using global feature from the first part and local feature together.

5.2. Second Experiment

In second part of experiment, three machine learning models that are Naïve Bayes, k-Nearest Neighbor, and Resilient propagation training algorithm are selected to compare with multilayer perceptrons with Levenberg-Marquardt training. Moreover, machine learning using only global features and using global and local features are also considered and evaluated in this experiment.

For Naïve Bayes model, the proportion of training and testing was 75:25. Naïve Bayes classifies only global feature and hybrid features in this part.

For k-Nearest Neighbor, in this part, this classifier is set as the first part that k is 1. Besides in this part, the k-Nearest Neighbor using only global features and that using hybrid features are analyzed.

For the Resilient propagation, this model is used to compare with our method. Five-fold cross validation was also considered together with 75:25 proportions of training images and testing images. The learning rate and the maximum epoch are set to be 0.25 and 100, respectively. The values of two classes are set to be '-1' and '1'. The linear activation function is chosen with two hidden layers consisting of three and four neurons,

respectively. All parameters are defined to yield the best results as possible. The image is assigned to the class whose output is greater than zero.

For our proposed method, neural network with the Levenberg-Marquardt training algorithm, dataset was divided into training set and testing set with the proportion of 75:25 as well as the five-fold cross validation. The parameters were set as described in chapter IV. The activation function is symmetrical linear function and the targets are set to be -1 and 1.

After the images were classified with provided models, the accuracy rate and False Acceptance Rate (FAR) are calculated to comparison, as shown in Table 5.2.

Table 5.2 : Result table of second part of experiment

Model	Feature	Accuracy rate(%)	FAR(%)
Lavenberg Marquardt	**Mixed**	**93.6**	**0**
Lavenberg Marquardt	Global	91.2	0.87
Naïve Bayes	Mixed	66.4	33.6
Naïve Bayes	Global	66.4	33.6
k-Nearest Neighbor	Mixed	76.6316	23.3684
k-Nearest Neighbor	Global	76.6316	23.3684
R-Prop	Mixed	81.6	0
R-Prop	Global	82.4	0.97

From the experimental result, among all models with only global features, it is obvious that multilayer perceptron with Levenberg-Marquardt training algorithm is better than other models in terms of accuracy and FAR.

CHAPTER VI

DISCUSSIONS

In this chapter, the discussions are separated into two parts as the experiment. Each part will be discussed each part of the experiment respectively.

From the first part of experimental results with the comparison among the methods, principal components gave lower accuracy than singular values because the PCA is based on eigenvector, the dimensionality reduction discarded the dimensions containing low variance. This means some information of feature vectors are lost whilst SVD decomposes the teeth image and the diagonal entries were used without any information loss. Additionally, this singular values gave higher accuracy, if they were combined with color histogram because the teeth from individuals also had the distinction in brightness and contrast. With the comparison of the gray histogram and color histogram, the color histogram gave better results. It means that gray images among many persons seem alike more than using color images. On the other hand, the personal behavior can change teeth color from time to time and the color can converge to that of other persons. Even though only the color histogram itself gave the high accuracy with small false acceptance, only histogram cannot guarantee the correctness of the result.

The results for the Naïve Bayes and the k-Nearest Neighbor were in the same direction that SVD feature was better than PCA. On the other hand, the result of color histogram was better than the SVD and color histogram. Although this gave high result but if the dataset is larger, this rate will decrease because it is possible that more than two persons have similar color of teeth. For resilient backpropagtion algorithm, accuracy of color histogram is higher than that of singular values concatenated with color histogram but false acceptance rate is also higher. For the failure identification, our system cannot handle blurry image since the blurry factor affected the deviate of color histogram and singular values.

From the second part of the experiment, Naive Bayes, k-Nearest Neighbor and R-Prop are not appropriate to the mixed features. The results prove that although adding local features a input of their systems, the performance of them is not high enough to practically use. In case of Naive Bayes and k-Nearest Neighbor, their accuracy rate is not even changed. This means that the local features does not affect the performance of those models. However, the performance of multilayer perceptrons with Levenberg-Marquardt algorithm is higher after adding than that in recent experiment because the number of person (classes) is increased. However, with the combination of global features and local features, the Levenberg-Marquardt backpropagation neural network can predict class of input with high accuracy. The FAR of the network is zero as well. Moreover, the failure in classification might be obtained from the physical characteristic of image when taken a shot. For instance, the motion blur causes the misclassified as shown in Figure 6.1.

Figure 6.1. Example of a motion blur image in dataset

CHAPTER VII

CONCLUSION

According from the comparison of results and discussion in the first part of the experiment, the features based on SVD and color histogram and multilayer perceptron with backpropagation learning and Levenberg-Marquardt algorithm were suitable for teeth recognition system. From this experiment, the proposed system gave the accuracy about 96% without any false acceptance. This means the SVD and color histogram feature can distinct each personal teeth-image. With the multilayer perceptron, backpropagation Levenberg-Marquardt learning algorithm, it can classify the features effectively without misclassifying wrong person. This system can be used for biometric identification with easy and not consume too much resource for fitting in small device.

For the second part of the experiment, according to the experiment results, the features based on SVD, color histogram and normalized teeth width with Levenberg-Marquardt training algorithm are suitable to be form teeth recognition system. From the result, our proposed system gave the accuracy rate about 93.6% without false acceptance. This means that the dataset is larger, the more specific features is needed to make all feature vectors easier to classify. However, selecting classification model is also significant. Finally, this proposed system can be used for biometric identification with easiness and low resource consumption.

It can be concluded that the result shows that the multilayer perceptron with Levenberg-Marquardt algorithm is satisfied as classification model with the linear transfer function. The feature vector that will be fed to the neural network is reduced dimension by using singular value decomposition. However the color histogram also assists to gain more accuracy. To yield better results, the teeth-width ratio is considered as local feature that can help the neural network performs better than the first part of the experiment using global features.

REFERNCES

1. Kiattisin, S. and Nadee, C. 2008.Improved PCA and LDA-based personal identification method. *Proceedings of 3rd International Symposium on Biomedial Engineering* Bangkok, Thailand : pp. 461-465.
2. Poon, B., Amin, M. A. and Yan, H. 2009. PCA based face recognition and testing criteria. *Proceeding of 8th International Conference on Machine Learning and Cybernetrics* Baoding, China : pp. 2945-2949.
3. Meedeniya, D.A. and Ratnaweera, D.A.A.C. 2007. Enhanced face recognition through variation of principle component analysis (PCA). *Proceeding of 2nd International Conference on Industrial and Information System* Sri Lanka : pp. 347-351.
4. Kim, D., Shin, J. and Hong, K. 2009. Teeth recognition based on multiple attempts in mobile device. *Journal of Network and Computer Applications* Vol. 33 : pp. 283-292.
5. Joen, J., Kim, J., Yoon, J. and Hong, K. 2008. Performance evaluation of teeth image recognition system based on Difference Image Entropy. *Proceeding of 3rd International Conference on Convergence and Hybrid Information Technology* : pp.967-972.
6. Kim, D. and Hong, K. 2008. Multimodal biometric authentication using teeth image and voice in mobile environment. *IEEE Transactions on Consumer Electronics* Vol. 54, No. 4 : pp. 1790-1797.
7. Veepraprasit, S. and Phimoltares, S. 2010. Neural network-based teeth recognition using singular value decomposition and color histogram. *Proceeding of 2nd International Conference on Information Engineering and Computer Science* China. Vol.3 : pp. 1311-1314
8. Veepraprasit, S. and Phimoltares, S. 2011. Hybrid Feature-Based Teeth Recognition system. *Proceeding of the 2011 IEEE International Conference on Imaging Systems Techniques* Penang, Malaysia. In press.
9. Jain, A.K., *Fundamentals of Digital Image Processing*. (Englewood Cliffs: NJ:

Prentice Hall, 1989) : pp. 150-153.

10. Riedmiller, M. and Braun, H.1993. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceeding of International Conference on Neural Networks* San Francisco : pp.586-591.
11. Nixon, M.S. and Aguado, A.S., *Features Extracton and Image Processing*. (Oxford: Great Britain: Newnes, 2002) : pp. 31-66.
12. Hagan, M.T., Demuth, H.B. and Beale, M.H., *Neural Network Design* (Boston, MA: PWS Publishing, 1996).
13. Nadee, C., Kumhom, P. and Chamnongthai, K. 2005.Improved PCA-Based Personal Identification Method Using Invariance Moment. *Proceedings of 3rd International Conference on Intelligent Sensing and Information Processing* Bengalore, India : pp. 243-247.
14. He, Y., Zhang, J. and He, X. 2010. The Acquisition of Tooth Profile Curve Based on Image Recognition. *Proceedings of International Conference on Measuring Technology and Mechatronic Automation* Changsha, China : pp. 742-745.
15. Faruqe, M.O. and Al Mehedi Hasan, M. 2009. Face Recognition Using PCA and SVM. *Proceedings of 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication* Hong Kong : pp. 97-101.
16. Hagan, M.T. and Menhaj, M.B. 1994. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks* Vol. 5, No. 6 : pp. 989-993.
17. Mitchell, T.M., *Machine Learning*. (USA: McGraw-Hill Science, 1997) : pp. 31-66.
18. Chomdej, T. and Pankaow, W., "Design and development of Dental Identification System in Forensic Medicine," (Theses of Master degree, Chulalongkorn University Medical Journal, 2005).
19. Shang, J., Zheng, X. and Zhang, Y. 2010. A Teeth Identification Method based on Fuzzy Recognition. *Proceedings of 2nd International Conference on Intelligent Human-Machine Systems and Cybernatics* Nanjing, China : pp. 271-275.
20. Renkjumnong, W., "SVD and PCA in image processing," (Mathematics Theses, Department of Mathematics and Statistics, Georgia State University, 2007).

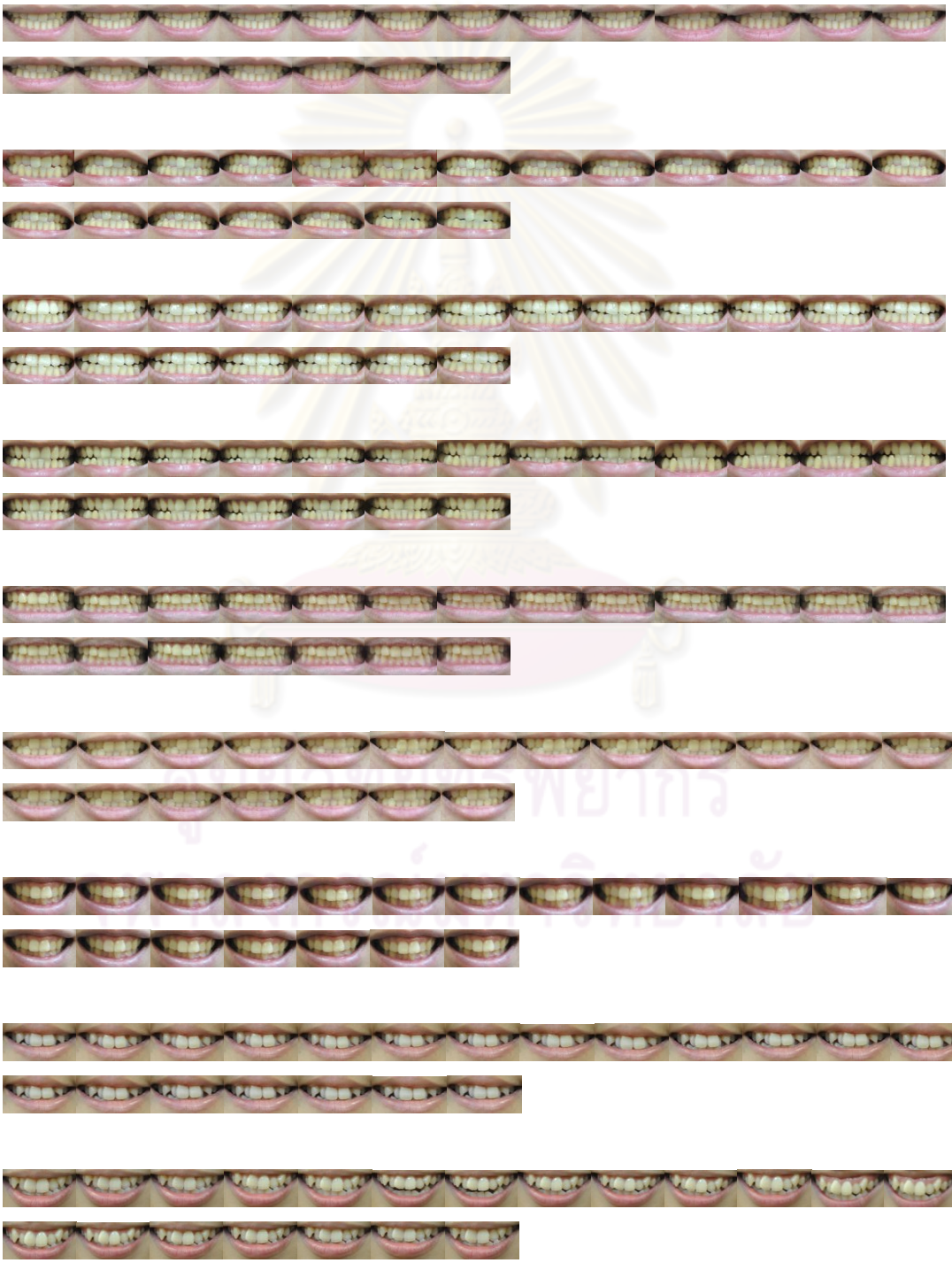
21. Turk, M.A. and Pentland, A.P. 1991. Face Recognition Using Eigenfaces. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Maui Hi, USA : pp. 586-591.

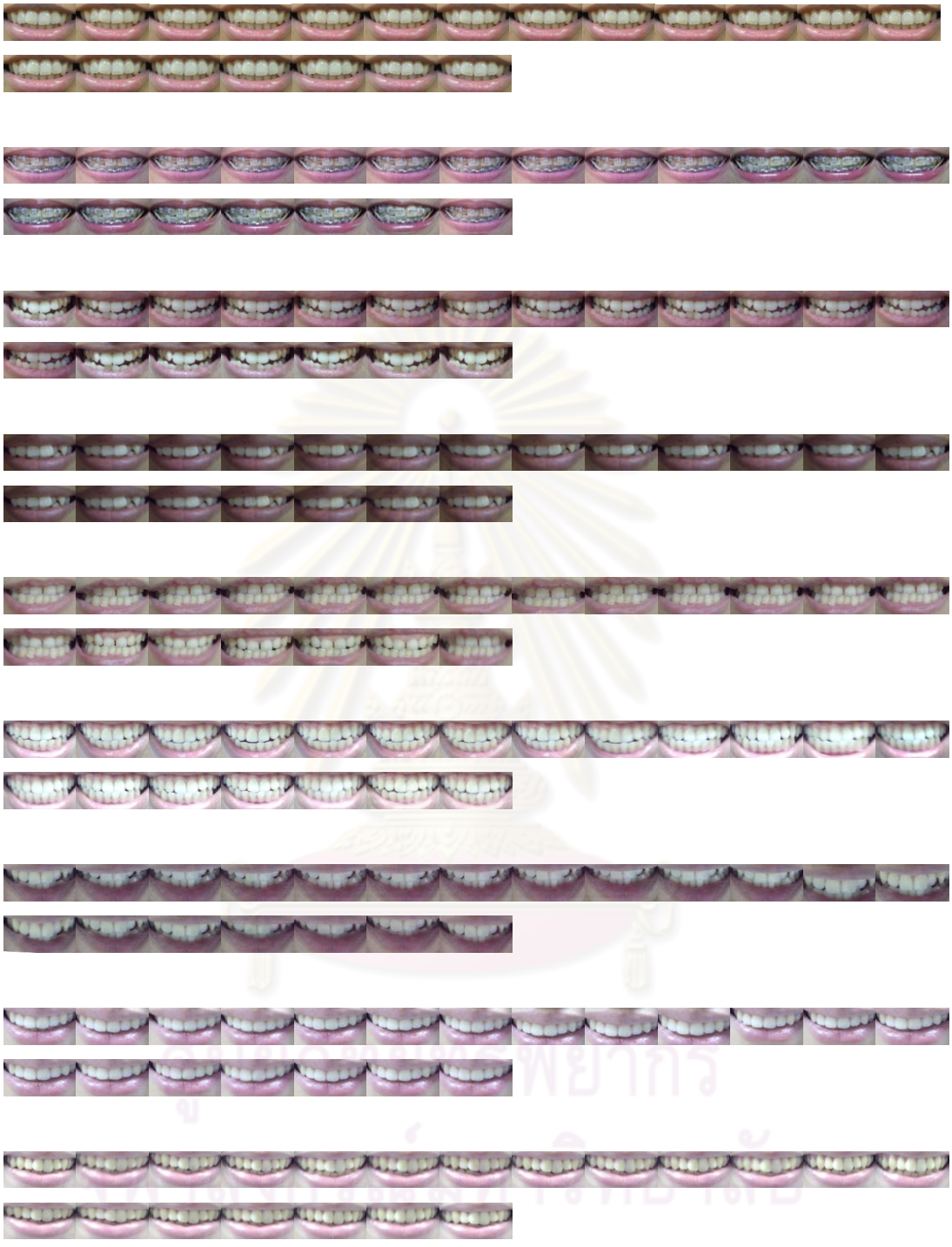
22. Will, T. Introduction to Singular Value Decomposition[online] November 3, 2003. http://www.uwlax.edu/faculty/will/svd/index.html

APPENDIX

THE PICTURE IN THE DATABASE

In this research, the example of picture from the entire the data is illustrated below separated in group followed each person. Each person was taken 20 pictures. There are 25 persons that give support to this research.





BIOGRAPHY

Suprachaya Veeraprasit was born at Bangkok, Thailand. She received a bachelor degree from Information and Communication Technology Program, Mahidol University. Now she is studying a Master Degree in Computer Science from Chulalongkorn University, and planning to qualify a doctorate degree in Computer Science too.