

การทำนายภาพหน้าคนโดยใช้โครงข่ายประสาทเทียมจากข้อมูลภาพและเนื้อหาบนเว็บเพจ



นางสาวชุติมน จิตติพรวณิช

ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ ภาควิชาคณิตศาสตร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

PREDICTION OF A HUMAN FACIAL IMAGE BY ANN USING IMAGE DATA AND ITS  
CONTENT ON WEB PAGES



Miss Chutimon Thitipornvanid

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 2009

Copyright of Chulalongkorn University

521749

Thesis Title PREDICTION OF A HUMAN FACIAL IMAGE BY ANN USING  
IMAGE DATA AND ITS CONTENT ON WEB PAGES


By Chutimon Thitipornvanid

Field of Study Computer Science and Information

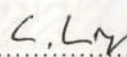
Thesis Advisor Siripun Sanguansintukul, Ph.D.

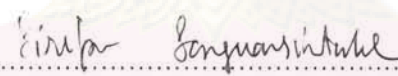
---


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  
(Professor Chidchanok Lursinsap, Ph.D.)

  
..... Thesis Advisor  
(Siripun Sanguansintukul, Ph.D.)

  
..... External Examiner  
(Worast Choochaiwattana, Ph.D.)

ศูนย์วิทยุโทรพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

ชุติมาน ฐิติพรวณิช : การทำนายภาพหน้าคนโดยใช้โครงข่ายประสาทเทียมจากข้อมูลภาพและเนื้อหาบนเว็บเพจ. (PREDICTION OF A HUMAN FACIAL IMAGE BY ANN USING IMAGE DATA AND ITS CONTENT ON WEB PAGES) อ.ที่  
 ปรักษาวิทยานิพนธ์หลัก: ดร. สิริพันธ์ สงวนสินธุกุล, 49 หน้า.

การเลือกใช้เมตาเดตาที่เหมาะสมเป็นสิ่งวิกฤต ในขณะที่สารสนเทศที่ดีและเหมาะสมของข้อมูลรูปภาพได้ให้ประโยชน์ช่วยลดความยุ่งยากของผู้ใช้และทำให้เกิดความสะดวกในการสืบค้นคัดแยกประเภทข้อมูลรูปภาพที่ต้องการออกจากกลุ่มรูปภาพจำนวนมากให้แก่ผู้ใช้ การเพิ่มคุณค่าให้กับรูปภาพไม่เพียงแต่แค่เพิ่มคุณค่าตรงส่วนตัวรูปภาพเท่านั้นแต่รวมไปถึงเทคนิคในการสืบค้น การวิจัยนี้ได้นำเสนอเทคนิคที่ง่ายแต่มีประสิทธิภาพในการทำนายภาพหน้าคนจากเว็บไซต์โดยใช้ข้อมูลพื้นฐานของรูปภาพและข้อมูลเนื้อหาที่อธิบายรูปภาพที่ปรากฏอยู่ในหน้าเว็บเพจ โดยผลทดลองของการวิจัยนี้ได้ให้ความถูกต้องในการทำนายภาพหน้าคนสูงถึง95% เทคนิคนี้อาจจะนำมาช่วยในงานห้องสมุด งานวิจัย และงานอื่นๆ สำหรับการประมวลผลอัตโนมัติและมีประสิทธิภาพในการแยกหมวดหมู่ของรูปหน้าคนออกจากกลุ่มของรูปภาพที่มากมาย

# ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา คณิตศาสตร์.....  
 สาขาวิชา วิทยาการคอมพิวเตอร์และสารสนเทศ  
 ปีการศึกษา 2552.....

ลายมือชื่อนิสิต ชุติมาน ฐิติพรวณิช  
 ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก.....  
 Siripon Senguanantikul

## 5073606823 : MAJOR COMPUTER SCIENCE AND INFORMATION

KEYWORDS : METADATA / PREDICTION / MULTI-LAYER PERCEPTRON / HUMAN FACIAL IMAGE / IMAGE MINING

CHUTIMON THITIPORNVNID : PREDICTION OF A HUMAN FACIAL IMAGE BY ANN USING IMAGE DATA AND ITS CONTENT ON WEB PAGES. THESIS  
ADVISOR : SIRIPUN SANGUANSINTUKUL, Ph.D., 49 pp.

Choosing the right metadata is critical, as good information (metadata) attached to an image will facilitate its visibility from a pile of other images. The image's value is enhanced not only by the quality of attached metadata but also by the technique of the search. This study proposes a technique that is simple but efficient to predict a single human image from a website using the basic image data and the embedded metadata of the image's content appearing on web pages. The result is very encouraging with the prediction accuracy of 95%. This technique may become a great assist to librarians, researchers and many others for automatically and efficiently identifying a set of human images out of a greater set of images.

Department : Mathematics

Student's Signature *ชุติมอน ฐิตทิพอรณนิธ*

Field of Study : .....

Advisor's Signature *Siripun Sanguansintukul*

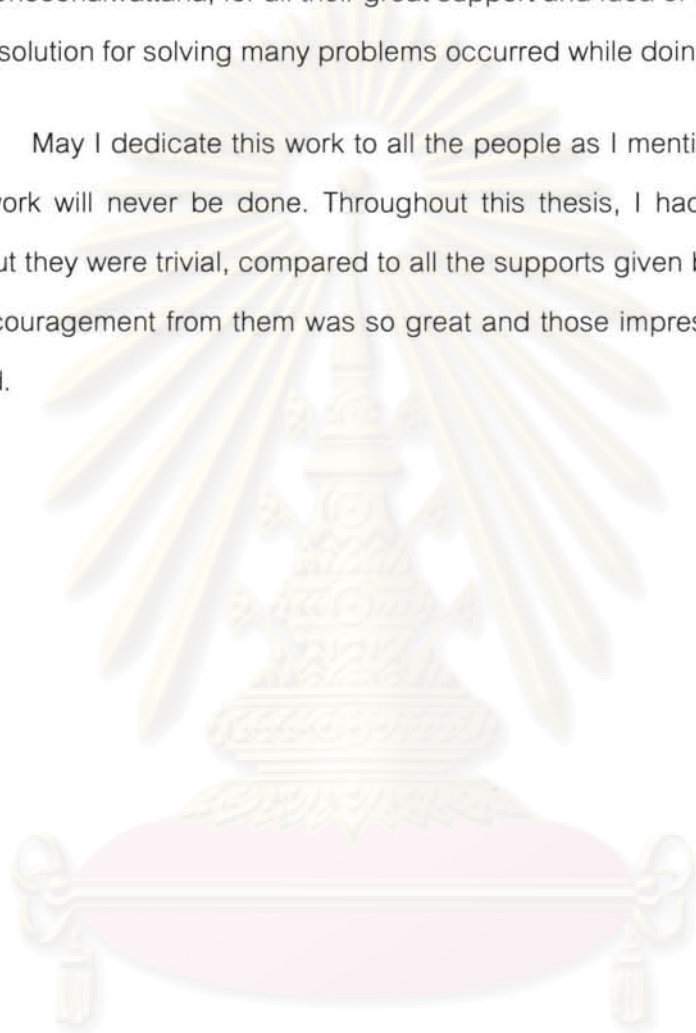
..... Computer Science and Information

Academic Year : 2009

## Acknowledgements

I would like to acknowledge my advisor, Dr. Siripun Sanguansintukul and Dr. Worasit Choochaiwattana, for all their great support and idea of this thesis. They also suggest the solution for solving many problems occurred while doing an experiment.

May I dedicate this work to all the people as I mentioned above. Without them, this work will never be done. Throughout this thesis, I had encountered many problems, but they were trivial, compared to all the supports given by these people. The warmest encouragement from them was so great and those impressions will always be remembered.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## Contents

	Page
Abstract (Thai).....	IV
Abstract (English).....	V
Acknowledgements.....	VI
Contents.....	VII
List of Tables.....	IX
List of Figures.....	X
Chapter	
1 Introduction.....	1
1.1 Problem Identification.....	2
1.2 Research Objectives.....	2
1.3 Scope.....	2
1.4 Research Methodology.....	3
1.5 Benefits.....	4
2 Fundamental Knowledge and Literature Review.....	5
2.1 Color Image Segmentation.....	5
2.1.1 Region-based approach.....	6
2.1.2 Pixel-based approach.....	7
2.1.3 Edge-based approach.....	8
2.1.4 Object Color specification.....	8
2.2 Color Space for Image Segmentation.....	9
2.2.1 The RGB color model.....	10
2.2.2 The CMY color model.....	12
2.2.3 The HIS, HSV and HLS color model.....	12
2.2.4 The YUV, YIQ and YCbCr color model.....	13
2.3 Neural Networks.....	14

2.3.1	Introduction to Neural Networks.....	14
2.3.2	Neural Networks architecture.....	15
2.4	Learning Algorithm.....	16
2.4.1	Backpropagation Algorithm.....	16
2.4.2	Task for Neural Networks.....	18
3	Experimental Application.....	21
3.1	Web page Metadata.....	21
3.2	Description of the system.....	21
3.3	Experimental Data Detail.....	23
3.3.1	Data Extraction.....	23
3.3.2	Low level data extraction.....	23
3.3.3	High level data extraction .....	26
3.3.4	Stage two: Training the network.....	31
3.3.4.1	Input Data.....	31
3.3.4.2	Output Data.....	31
4	Experimental and Results.....	34
4.1	Experimental results.....	34
4.1.1	Experiment on Images using two color models: RGB and YCbCr....	34
4.1.2	Experiment with and without text surrounding images.....	36
5	Discussion and Future works.....	39
5.1	Conclusions.....	39
5.2	Discussions.....	40
5.3	Future works.....	40
	References.....	42
	Appendix A.....	46
	Vita.....	49



## List of Tables

Table		Page
1.1	Research methodology time table.....	4
3.1	A sample of high level extraction.....	30
4.1	Human facial image classification results using RGB.....	34
4.2	The face classification result using YCbCr.....	35
4.3	The face classification result using YCbCr with text information (high level description).....	37
4.4	The face classification result using YCbCr without text information (high level description).....	37



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## List of Figures

Figure		Page
2.1	Examples of color images from CNN.....	5
2.2	A schematic of the RGB color cube.....	11
2.3	RGB 24-bit color cube .....	11
2.4	A range of hues.....	12
2.5	Two layers fully interconnected neural network.....	15
2.6	A typical Multiple Layer Perceptron (MLP) architecture.....	16
3.1	The experimental procedure.....	22
3.2	A Simple of human facial image using RGB.....	24
3.3	A Simple of human facial image using YCbCr.....	26
3.4	Website dictionary corpus.....	27
3.5	An example of file name.....	28
3.6	An example of caption name.....	29
3.7	An example of title name .....	30
3.8	A typical Multilayer Perceptron ANNs Architecture.....	32
4.1	An example of detecting human skin tone using RGB and YCbCr model.....	36
5.1	The experimental process.....	39

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

# CHAPTER I

## INTRODUCTION

Over the last few years, the interest in the digital image has grown rapidly on the World Wide Web visual and audio resources in the form of still pictures, graphics, audio, speech, and video play an increasingly pervasive role in our lives. Images, especially, are rich in information content. However, their contents are also complex. Metadata can describe image data. Creating an image metadata is difficult, subjective, time consuming and expensive.

Typically, metadata is defined as "information about information" or "information to describe other information". According to Taylor [1], metadata are referred to as a description of attributes and contents of an information package that may include descriptive information about the content, quality and condition, or characteristics of data. The objective of the metadata is to find/locate, identify, select, obtain, and navigate [3]. Metadata are used to expedite and enhance searching for resources. Metadata become important on the WWW due to the need to find useful information from the much larger available information.

Metadata in digital libraries can be divided into 3 categories: descriptive, structural, and administrative. Descriptive metadata is information derived from the content of the data. Elements include: title, name, edition and publication date. Structural metadata is related to information about the structure, format and composition of the data. Administrative metadata are inherently extrinsic properties such as who, what, why, where of the object's creation and management. Metadata is not limited to documents. Any resources such as video, audio including image can be described with an appropriate metadata element set.

This information is increasingly available to the public in the electronic form. Photographs are captured and posted for various purposes. Specifically, the image of human face becomes significant on different activities such as face identification, face recognition and face tracking [2]. Therefore, an image searching system to enhance information retrieval on human images is expected to become of great interest.

The central focus of this study is to develop a simple but efficient technique to classify a single human image from a website using basic image data and the image content appearing on web pages. This technique may become a great assist to the librarians, researchers and many others for automatically and efficiently identifying a set of human images out of a greater set of images.

### 1.1 Problem Identification

From the growth of electronic information described above, numerous problems arise in the images on the World Wide Web. Some of the problems are:

1. Choosing the right metadata that attached to an image is critical to create a proposed model.
2. The propose model that is simple but efficient to predict a single human image on the websites.

### 1.2 Research Objectives

This research attempts to study the following aspects:

1. To classify a single human image from others images
2. To measure the performance of ANN in classifying human images on the web page.
3. To compare between RGB and YCbCr color model, that are implemented in the image processing program
4. To verify that metadata surrounding images can improve the classification performance

### 1.3 Scope

Due to the sheer volume images of the World Wide Web, the proposed model will confine the scope of study within the following domain of applications:

1. The propose model is focused on predicting only a single human color images not including gray-scale images.

2. Since CNN has been widely used in researches. A CNN website has been initially investigated for the experiments.

#### 1.4 Research Methodology

In order to achieve the above objectives, the following tasks will be carried out by means of the theoretical work described below:

1. Study concepts of related technologies.
  - Study basic concept of face detection technology.
  - Study basic concept of Metadata.
2. Define the problem statement.
3. Study algorithms to implement the RGB and YCbCr model.
4. Conduct an experiment to verify the viability of the proposed model.
5. Write the thesis

Below is a time table covered all of the above tasks.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



Table 1.1 : Research methodology time table

No	Tasks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	Investigate related technologies																			
2	Define the statement of the problem																			
3	An algorithm to construct the RGB and YCbCr model																			
4	Conduct an experiment to verify the viability of the proposed model.																			
5	Write the thesis																			

### 1.5 Benefits

The proposed method aims to classify single human color images on the webpage using the simple image information and the image contents.

## CHAPTER II

### FUNDAMENTAL KNOWLEDGE AND LITERATURE REVIEW

One of the interests in image analysis is extracting the meaningful information from images. The most well-known technique in the digital image processing is the image segmentation. For example, if we have a digital image of a human on the lawn background. A natural partitioning would be to have all brown pixels labeled as 'face' and the remaining green pixel labeled as 'ground'. In general, segmentation techniques can be categorized into four groups [3] region-based techniques, pixel-based techniques, edge-based techniques and model-based techniques. Figure 2.1 shows examples of color images: human face and non human face.



Figure 2.1 : Examples of color images from CNN

#### 2.1 Color Image Segmentation

The goal of image segmentation is the process to classify features which correspond to objects of interest from the background of the image. Color Image segmentation techniques attempts at grey-scale image segmentation based on three main categories: region-based, pixel-based and edge based techniques [4]. Region-based approaches try to find partitions of the image pixels into sets corresponding to coherent image properties such as brightness, color and texture. Contour-based approaches usually start with a first stage of edge detection, followed by a linking

process that seeks to exploit curvilinear continuity. Boundaries of regions can be defined to be contours.

### 2.1.1 Region-based approach

A region based technique for doing image classification focuses on continuity of a region in the image. The region based method attempt to group pixel into objects using image segmentation process based on a chosen similarity image partition e.g. texture, color, intensity and then use the spectral, spatial and contextual information inherent in these objects to classify the whole image.

Unlike the pixel-based techniques, region-based techniques consider both color distribution in color space and spatial constraints [5]. One of the strengths is the ability to extract real world objects, proper shape and accurate classification. It eliminates the mixed pixel problem, which is suffered by most pixel based methods. In general, the region-based approach includes the region growing in the first step which is the process of grouping neighboring pixels or collection of pixels of similar properties into larger region and then employing the following techniques :

- Merging algorithm: this algorithm will produce larger region, in which neighbor regions are compared and merged if they are close enough in some property such as a distance measure linked to color similarity.
- Splitting algorithm: this algorithm will split the image into smaller and smaller region, in which large non-uniform regions are broken up into smaller areas which may be uniform.

#### Region Merging

Merging must start from a uniform seed region. Some work has been done in discovering a suitable seed region. One method is to divide the image into  $2 \times 2$  or  $4 \times 4$  blocks and check each one. Another method is to divide the image into strips, and then subdivide the strips further. In the worst case the seed will be a single pixel. Once a seed has been found, its neighbors are merged until no more neighboring



regions conform to the uniformity criterion. At this point the region is extracted from the image, and a further seed is used to merge another region.

### **Region Splitting**

This algorithm starts from the whole image, and divides the image up until each sub region is uniform. The usual criterion for stopping the splitting process is when the properties of a newly split pair do not differ from those of the original region by more than a threshold.

The main problem with this type of algorithm is the difficulty in deciding where to make the partition. Early algorithms used some regular decomposition methods, and for some classes there are satisfactory, however, in most cases splitting is used as a first stage of a split/merge.

This algorithm combines the method of splitting and merging. In all cases some uniformity criterion must be applied to decide if a region should be split, or two regions should merged. This criterion is based on some region property which will be defined by the application, and could be one of many measurable image attributes such as mean intensity, color etc. Uniformity criteria can be defined by setting limits on the measured property, or by using statistical measures, such as standard deviation or variance.

Kenong Wu and Martin D.Levine described the solution to efficient region growing problem with multiview range images of a 3D object by using the region-based approach. They presented the technique for segmenting the object into parts at deep surface concavities. Region-based approaches aim to find image regions that respond to object surface patches, and then group these patches into individual parts based on particular surface configurations. The advantage of this classification approach is its ability to approximate the shape of an object.

#### **2.1.2 Pixel-based approach**

As discussed above example, color is one of the easiest ways to identify which pixels belong to which object. Pixel-based techniques do not consider the spatial context but only decide solely on the basis of the color features at each pixel in the

image, such as standard Markov random fields, consider image segmentation as a labeling issue at the pixel level.

The vector gradient operator employs the concept of a gradient operator on a three channel color vector space. Di Zenzo proposed a combination of three chromatic gradients for getting a global gradient. He implemented this operator in the RGB color space.

### 2.1.3 Edge-based approach

In many ways, edge detection can be considered as the dual of image segmentation. Instead of finding the regions associated with various objects, the goal of edges detection is to find the boundaries of interested objects. Once the edges have been found, the interior can be filled-in to obtain the region associated with an object. The primary hypothesis used in edge detection is that there is a change in pixel color or intensity at pixels on the boundary objects. Edge detection is relatively easy. For example, the color change from pixels on a face to pixels on a leaf or the lawn would be very large.

Once image segmentation has been performed, it is often possible to perform computer-based analysis of the position, size or shapes of objects in an image. There are two famous signal edge detectors: the Canny operator and the Shen-Castan (ISEF) method (Bach,1986; Shen & Castan 1992). The Canny algorithm convolves the image with the derivative of a Gaussian function and then performs non-maximum suppression and hysteresis threshold. The Shen-Castan algorithm convolves the image with the Infinite Symmetric Exponential Filter (ISEF), computes the binary Laplacian image, suppresses false zero crossings, performs adaptive gradient threshold, and also applies the threshold.

### 2.1.4 Object color specification

The color histogram [6] is the most common approach used for pixel-based technique to examine the distribution of colors within an image. It is a discrete approximation to the probability density function for colors in an image. A color



histogram  $h(c)$  is calculated by counting the number of pixels in the image  $I(x,y)$  with each distinct color  $c$ :

$$h(c) = \sum_{x,y} \begin{cases} 1 & \text{if } I(x,y) \cong c \\ 0 & \text{otherwise} \end{cases}$$

To normalize the image area it is common to divide each value of  $h(c)$  by the number of pixels in the image. The size of the color histogram depends on the number of colors in the input image and also the needs of the application using the color histogram. In the case where we are given an 8-bit pseudo-color image as input, the number of colors in the image has been drastically reduced via color quantization. It can represent the color histogram of such an image using an array of 256 counters.

When they are given a 24-bit RGB image with 8 bits for each of red, green and blue, a complete color histogram would require a  $256 * 256 * 256$  array of counters. This would require far more memory than is typically available, so it is common to discard the least significant bits of each color byte to reduce the size of the color histogram.

## 2.2 Color Space for Image Segmentation

From methods described above, the choice of using a proper color space is very important to the goal of the application. A color model (or color space) is a model specification created to explain, define, and specify color by a single point in a three-dimensional color coordinate system. In term of digital image processing, the hardware-oriented models most commonly used in practice are the RGB (red, green, blue) model for color monitors and a broad class of color video cameras; the CMY (cyan, magenta, yellow) and CMYK (cyan, magenta, yellow, black) models for color printing; and the HSI (hue, saturation, intensity) model, which corresponds closely to the way humans describe and interpret color.

### 2.2.1 The RGB color model

The RGB model is the unit cube subset of the 3D Cartesian coordinate system. This RGB color model is represented in three-dimensions which is commonly known in use today for digital images. The advantages of the RGB model are:

1. The RGB model is the most popular and widely used for pictures acquired by digital cameras and each color in RGB appears in its primary spectral components of red, green and blue. Thus, it is easy for programmers to understand and program.
2. The RGB model is used for television and the best-known application of this color model is the cathode-ray tube used in color televisions.

In the RGB color model, colors are represented by varying intensities of red, green and blue. As illustrated in Figure 2.2 and Figure 2.3, it is a RGB unit cube [7] subset of the 3D Cartesian coordinated system of the color subspace. RGB primary values are at three corners; the secondary colors cyan, magenta, and yellow are at three other corners; black is at the origin; and white is a combination of the three primary colors in the proper amounts at the corner farthest from the origin. All values of R, G, and B are assumed to be in the range  $[0,1]$ . As the figure shows, the monochrome (gray scale) vector stretches from black  $(0,0,0$  primaries) to white  $(1,1,1$  primaries). The color gamut covered by the RGB model is defined by the chromatic of the CRT (computer display) phosphors. Therefore, it follows that devices using different phosphors will have different color gamuts.

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

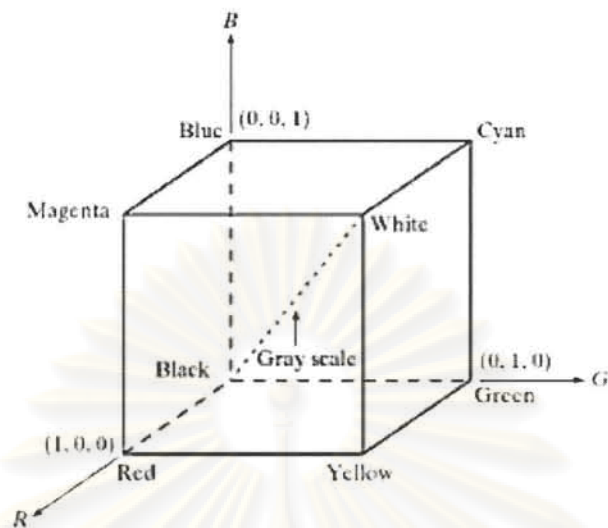


Figure 2.2 : A schematic of the RGB color cube [7]

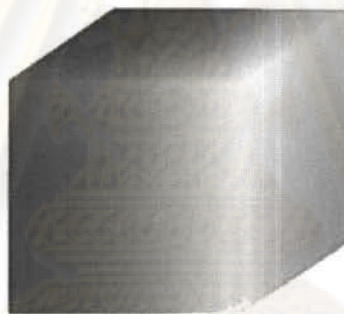


Figure 2.3 : RGB 24-bit color cube [7]

Images represented in the RGB color model consists of three component images, one for each primary color. The number of bits used to represent each pixel in RGB space is called the pixel depth. Consider an RGB image in which each of the red, green, and blue image is an 8-bit image. Under these conditions each RGB color pixel [that is, a triplet of values (R,G,B)] is said to have a depth of 24 bits is  $(2^8)^3 = 16,777,216$ .

Even though the RGB model is used for a number of years, a major drawback of the RGB space is that it is senseless[8]. For example, it is difficult for the user to understand or get a sense of what color R=100, G=50, and B=80 is and the difference between R=100,G=50,B=50 and R=100,G=150,B=150.

### 2.2.2 The CMY color model

The CMY color model represents the human perception of color more closely than the standard RGB model used in computer graphics hardware. In CMY, the three primary colors are cyan, magenta, and yellow are complements of red, green, and blue, respectively. The advantages of the CMY model are:

1. CMY model is mainly used for color printed media such as printers and copiers in printing industry.
2. CMY model is always used in the photography.

The subset of the Cartesian coordinate system for the CMY model is identical to RGB except that white is the origin, rather than black. The corresponding from RGB to CMY values [15] can be calculated by subtracting the RGB values from [1,1,1].

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The disadvantage of CMY model is same as the RGB.

### 2.2.3 The HIS, HSV and HLS color model

When humans view a color object, it describes by its hue, saturation, and brightness. HSI color model stands for Hue, Saturation, and Intensity [9]. Hue represents an actual wavelength of a color that describes a pure color's name such as pure red, green, yellow, orange, blue, and so on, whereas saturation gives a measure of the degree to which a pure color is diluted by white light. For instance, the color red is a 100% saturated color, but pink is a low saturation color due to the a mount of white in it. Intensity indicates the lightness of the color. It ranges from black to white. Figure 2.4 shows a range of hues.



Figure 2.4: A range of hues.



HLS (hue, lightness, and saturation) is similar to HIS; the term lightness is used rather than intensity. The difference between HIS and HSV (a color is represented using three components: hue (H), saturation (S), and value (V)) lies in the computation of the brightness component (I or V), which determines the distribution and dynamic range of both brightness (I or V) and saturation (S). HSV is the color selection model used most often in illustration and image programs, like Fireworks and Freehand. Color selection based on these criteria is often presented as a color wheel, with hues along the outer edge at full saturation, and with saturation decreasing as you move to the center of the circle. Value or intensity is adjusted with a brightness bar. Hue is presented as an angle point, while saturation and value are measured as a percentage between 0 and 100: The advantages of the HSV model are:

1. HSV color model is the most useful in painting or in drawing programs.
2. HSV color model is to be used more intuitive in manipulating color and to approximate the way human perceive and interpret color. So users could choose the color they want easily by indicating the hue, saturation and intensity values independently.

#### 2.2.4 The YUV, YIQ and YCbCr color model

The YUV model is widely used in image compression and processing applications. Y represents the luminance of a color, while U and V represent the chromaticity of a color. The luminance (Y) component is separated from the chromatic components in this space. The YIQ color space is derived from the YUB color space. The I stands for In-phase and Q for Quadrature, which is the modulation method used to transmit the color information. YCbCr is a scale and offset version of the YUB color space. Those color spaces are difficult for users to deal with, because they do not directly refer to intuitive notions of hue, saturation and brightness.



## 2.3 Neural Networks

### 2.3.1 Introduction to Neural Networks

Neural networks have become the great interest since 1943 by McCulloch & Pitts [10]. The networks are suggested as system models to process information like the human brains. The key terms to explain a "neural network" are as follow [11]:

"A neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns"

In general, the term "network" will be used to refer to any system of artificial neural. The neural networks technology [12] performs a kind of automatic feature extraction. For example, the hidden layer nodes in a feed-forward network trained by backpropagation can be thought of as extracting features which will ultimately be resolved into a classification at the output layer. If there are multiple hidden layers, the hidden layer neurons in each successive layer extract features of increasing complexity train to and these features may or may not have desirable properties.

Most neural networks involve combination, activation, error, and objective functions.

**Combination functions:** Each non-input unit in a neural network combines values that are fed into it via synaptic connections from other units, producing a single value called the "net input". For the function that combines values, it is called the "combination function". The combination function is a vector-to-scalar function. Most Neural networks use either a linear combination function (as in MLPs) or a Euclidean distance combination function (as in RBF networks).

**Activation functions:** Most activation functions are also known as threshold functions or squashing functions. The units in neural networks transform their net inputs by using a scalar-to-scalar function called an "activation function", yielding a value called the unit's "activation". Except possibly for output units, the activation value

is fed via synaptic connections to one or more other units. The activation function is sometimes called a "transfer" function. An activation function with a bounded range is often called "squashing" functions, such as the commonly used tanh (hyperbolic tangent) and logistic ( $1/1+\exp(-x)$ ) function. If a unit does not transform its net input, it is said to have an "identity" or "linear" activation functions.

**Objective function:** The objective function is what you directly try to minimize during training. Neural network training is often performed by trying to minimize the total error or the average error for the training set. However, minimizing training error can lead to over-fitting and poor generalization if the number of training cases is small relative to the complexity of the network.

### 2.3.2 Neural Networks architecture

In the architecture of neural networks, several perceptrons may be grouped together to form a neural network where the two layers of neurons are fully interconnected, but there is no interconnection between neurons in the same layer. This result in a network as shown in Figure 2.5.

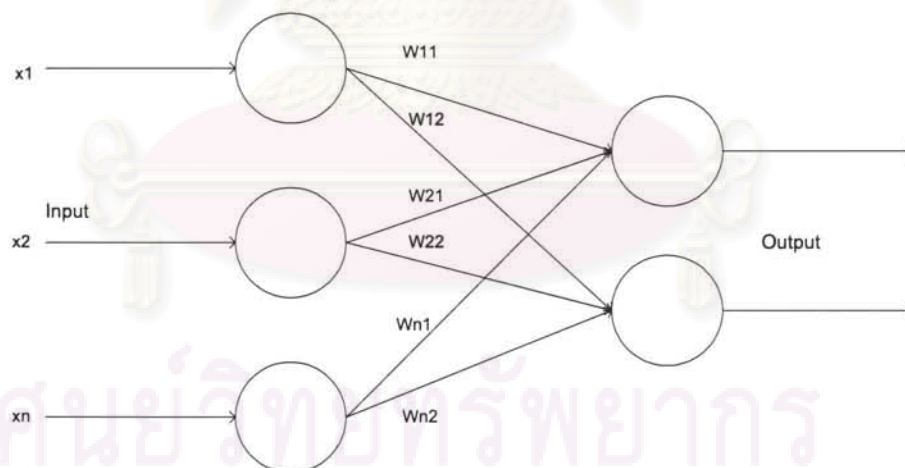


Figure 2.5: Two layers fully interconnected neural network.

Some common examples of different types of neural networks are Multiple Layer Perceptrons (MLP). The MLP network is a widely reported and used neural network. It consists of an input layer of neurons, one or more hidden layers of neurons, and an output layer of neurons as illustrated in the very simple structure of

Figure 2.6. Each neuron calculates the weighted sum of its inputs, and uses this sum as the input of an activation function, which is commonly a sigmoid function.

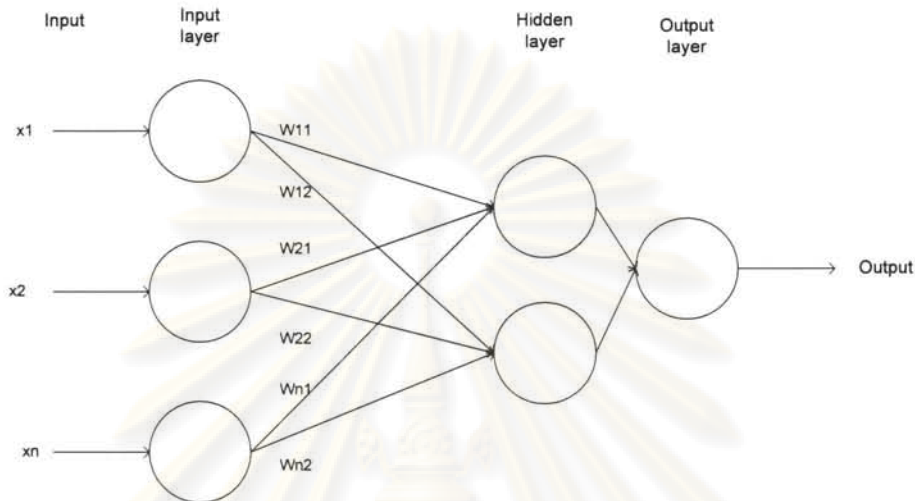


Figure 2.6: A typical Multiple Layer Perceptron (MLP) architecture.

## 2.4 Neural Networks architecture

Neural networks are trained by two main types of learning algorithms [13]: supervised and unsupervised learning algorithm.

**Supervised learning:** A supervised learning algorithm adjusts the strengths or weights of the inter-neuron connections according to the difference between the desired and actual network outputs corresponding to a given input. Thus, supervised learning requires a teacher or supervisor to provide desired or target output signals.

**Unsupervised learning:** Unsupervised learning algorithms do not require the desired outputs to be known. During training, only input patterns are presented to the neural network which automatically adapts the weights of its connections to cluster the input patterns into groups with similar features.

### 2.4.1 Backpropagation Algorithm

The backpropagation algorithm trains a given feed-forward multiplayer neural network for a given set of input patterns with known classifications. When each entry of the sample set is presented to the network, the network examines its output



response to the sample input pattern. The output response is then compared to the known and desired output and the error value is calculated. Based on the error, the connection weights are adjusted. The backpropagation algorithm is based on Widrow-Hoff delta learning rule. The backpropagation algorithm outline are based on follows[14]:

1. Initialize all the weight  $w_{ij}^l$  to small random values (typically between  $-0.1$  and  $+0.1$ ), where  $w_{ij}^l$  is the weight in layer  $l$  which connects unit  $i$  in layer  $l-1$  with unit  $j$  in layer  $l$ .

2. Initialize the activations for the threshold units. The values of these units will never change.

$$x_0 = 1, h_0^l = 1 \text{ for } 1 \leq l \leq L.$$

3. Choose a pattern  $P_k$  where  $P_k = (p_1, p_2, \dots, p_n)$  and apply it to the input layer so that:

$$x_i = p_i \text{ for } 1 \leq i \leq n.$$

4. Propagate the signal forwards through the network, using for the first layer:

$$h_j^l = \frac{1}{1 + e^{-NET_j^l}} \text{ for } 1 \leq j \leq H_l \text{ and where } NET_j^l = \sum_{i=0}^n w_{ij}^l x_i$$

For layer  $1 \leq l \leq L$ :

$$h_j^l = \frac{1}{1 + e^{-NET_j^l}} \text{ For layer } 1 \leq l \leq L:$$

$$h_j^l = \frac{1}{1 + e^{-NET_j^l}} \text{ for } 1 \leq j \leq H_l \text{ and where } NET_j^l = \sum_{i=0}^{H_{l-1}} w_{ij}^l h_i^{l-1}$$

For the output layer:

$$y = \frac{1}{1 + e^{-NET_l^L}} \text{ and where } NET_l^L = \sum_{i=0}^{H_{l-1}} w_{il}^L h_i^{L-1}$$

Where  $H_l$  is the number of units in the hidden layer  $l$ .

5. Compute the error for the output layer with:

$$\delta_l^L = (y' - y)y(1 - y)$$



by comparing the actual output  $y$  with the desired one  $y'$  for pattern  $P_k$  that is being considered.

6. Compute the errors for the preceding layers by propagating the errors backwards. For hidden layer L-1:

$$\delta_i^{(L-1)} = h_i^{(L-1)}(1 - h_i^{(L-1)})\delta_i^L w_{il}^L \text{ for } l \leq i \leq H_{L-1}$$

For hidden layer  $L-1 > l \geq 1$ :

$$\delta_i^l = h_i^l(1 - h_i^l) \sum_{j=l}^{H_{l+1}} \delta_j^{l+1} w_{ij}^{l+1} \text{ for } l \leq i \leq H_l$$

7. Update weights with the following formula:

$$w_{ij}^l = w_{ij}^l + \Delta w_{ij}^l$$

Where  $\Delta w_{ij}^l$  for the first hidden layer:

$$\Delta w_{ij}^l = \eta \delta_j^l x_i \text{ for } 0 \leq i \leq n \text{ and } l \leq j \leq H_l$$

For weights between two layers of hidden units:

$$\Delta w_{ij}^l = \eta \delta_i^L h_i^{L-1} \text{ for } 0 \leq i \leq H_{L-1}, l \leq j \leq H_l \text{ and } 1 \leq l \leq L$$

For weights between the last hidden layer and the output layer:

$$\Delta w_{ij}^l = \eta \delta_i^L h_i^{L-1} \text{ for } 0 \leq i \leq H_{L-1}$$

8. Go back to step 3 and repeat for the next pattern.

#### 2.4.2 Task for Neural Networks

Artificial neural networks are viable and important computational models for wide variety of problems. It applicable in virtually every situation in which the relationship between the predictor variables (independents, inputs) and predicted variables (dependents, outputs) exists, even when that relationship is very complex and not easy to articulate in the usual terms of "corrections" or "differences between groups." There are a lot of applications that can benefit from neural network and can be grouped in following categories:

- **Clustering:**

A clustering algorithm explores the similarity between patterns and places similar patterns in a cluster. Best known applications include clustering groups of bank loan applicants.

- **Classification/Pattern recognition:**

The task of classifier is operation of the pattern (data) recognition system. This is an important application of neural networks. An input pattern is described using a variety of features. Neural networks are often used to solve such complex classification problems based on prior knowledge of the patterns. This category can be implemented by using a feed-forward neural network. During training, the network is trained to associated outputs with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern. The power of neural networks comes to life when a pattern that has no output associated with it, is given as an input. In this case, the network gives the output that corresponds to a taught input pattern that is least different from the given pattern.

Problem has been applied successful in detection of medical phenomena. A variety of health-related indices (e.g., a combination of heart rate, levels of various substances in the blood, respiration rate) can be monitored. Neural networks have been used to recognize this predictive pattern so that the appropriate treatment can be prescribed [15].

- **Function approximation:**

The tasks of function approximation is to find an estimate of the unknown function  $f()$  subject to noise. Various engineering and scientific disciplines require function approximation.

- **Prediction:**

The task is to forecast some future values of a time-sequenced data. Prediction has a significant impact on decision support systems. Prediction differs from Function approximation by considering time factor.

For example, problem has been applied in the prediction for stock market forecasting [16]. Neural networks are being used by many technical analysts to make predictions about stock prices based upon a large number of factors such as past performance of other stocks and various economic indicators.

Examples of problems that neural networks have been applied successfully are:

- Engine management. Neural networks have been used to analyze the input of sensors from an engine. The neural network controls the various parameters within which the engine functions, in order to achieve a particular goal, such as minimizing fuel consumption.
- Monitoring the condition of machinery. Neural networks can be instrumental in cutting costs by bringing additional expertise to scheduling the preventive maintenance of machines. A Neural network can be trained to distinguish between the sounds a machine makes when it is running normally ("false alarms") versus when it is on the verge of a problem. After this training period, the expertise of the network can be used to warn a technician of an upcoming breakdown, before it occurs and causes costly unforeseen "downtime."



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## CHAPTER III

### EXPERIMENTAL APPLICATION

#### 3.1 Web page Metadata

Generally, resources on the web pages contain various image types such as content, logo, icon and image. One of the most valuable features that offers the web site owner to control their websites is META information. Metadata has the advantage for resource discovery. The META element provides metadata such as the document's keywords, description, and author. So, Metadata can provide its description with the link to the web document. Generally, the metadata of image resources published on the Web provide the meaning of the image. Image resources associated to metadata, for example, file name, caption name and title name are among the simplest and useful forms of metadata that allows people to understand more about images.

Therefore, this study endeavors to utilize the metadata information and the basic image information to predict a single facial human image on the web pages. This technique may become a great assist to the librarians, researchers and many others to automatically and efficiently identify a set of human images out of a greater set of images.

#### 3.2 Description of the system

The experiment procedure has been developed to classify a single human facial image. The experiment procedure can be divided into three stages: In the first stage, interested information from the image is extracted. The information includes the low level and high level data. The low level data consists of width, height, size and orientation of the images pixels. Whereas, the high level data are composed of file name, caption name, title name and position name which are extracted from the tags contents surrounding the image. In the second stage, both low and high level data will be used to train the network to classify the human facial image. Finally, the trained network model will be evaluated for its performance. The summary of the process is as follows:



A. Data extraction;

3.2.1 Low level data extraction

3.2.2 High level data extraction

B. Training the network;

C. Testing the performance of model from the network;

The procedure in the experiment can be illustrated as in Figure. 3.1.

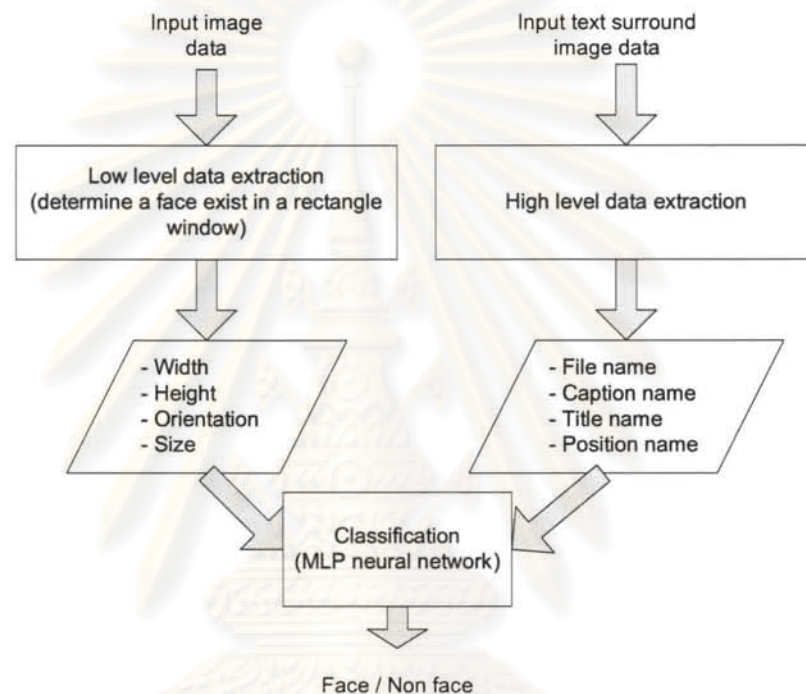


Figure 3.1 : The experimental procedure.

Figure 3.1 shows an overview of the experimental procedure. The input data in the system is obtained from 2 sources: image data and text surrounding the image data. In part of the low level extraction which is display on the left-hand, the input images in the JPEG format are loaded into the program [17]. The program tried to detect the human facial image, then a rectangle window is drawn on the detected human face. On the right-hand side, texts surrounding image of meta information are extracted. The information from the low and high level extractions are then feed to the network for training. All experimental images were obtained from the CNN website. The CNN website has been chosen since it provides various kinds of images and it is one of the popular website. Figure 3.3 shows an example of a rectangle window on the human

face. It is important to say that images color models applied in the experiment is YCbCr model. However, RGB model will also be employed for the performance evaluation purpose. The discussion about the experimental detail will be in Section 3.3.

### 3.3 Experimental Data Detail

#### 3.3.1 Data Extraction

The objective of this stage is to automatically extract the basic information of the images, together with, the contents surrounding the images from the website. Here, a program called crawler was developed to collect the information from the CNN website. The crawler performed the following tasks:

1. Download images:

The objective of this step is to automatically collect images from the website. In this study, CNN website is the main focus. The program will download the image files in the form of JPEG format. Then, all images are stored in a specific folder on the computer. These images will be further processed in section 3.3.2.

2. Download contents:

The objective of this step is to automatically download text information surrounding images. The text information comes from 4 tags: <file name>, <title>, <caption> and <alt>. For tags information, it will be further processed in section 3.3.3.

#### 3.3.2 Low level data extraction

After the crawler collected images in JPEG format. Image files in JPEG will be converted to GIF format. Then, the low level extraction process is began. The objective of the low level extraction is to detect skin color which is the important feature of the human face. To define the skin color, this study employs two different skin color models: RGB and YCbCr. The skin color of human face will be detected using a rectangle window from the background. There are 4 attributes extracted from each image: width, height, size and orientation of the human face.

### The RGB color model

In the RGB color model, there are three normalized components  $r$ ,  $g$  and  $b$ , which are known as pure colors. The  $r$  is the red value,  $g$  is the green value and  $b$  is the blue value. In the low level extraction, the RGB color space is chosen to detect the skin color following the equations in the program [17] as below:

$$g_{rgb}(x, y) = \begin{cases} (r, g, b); & \text{if } (R(x, y) \in [120, 225], G(x, y) \in [100, 225], B(x, y) \\ & \in [100, 225]) \wedge (R(x, y) \neq G(x, y) \neq B(x, y)) \\ (0, 0, 0); & \text{otherwise} \end{cases} \quad [Equation 3.2]$$

Where  $g_{rgb}(x, y)$  is the result from threshold

$R(x, y)$  is the Red color value at pixel  $(x, y)$

$G(x, y)$  is the Green color value at pixel  $(x, y)$

$B(x, y)$  is the Blue color value at pixel  $(x, y)$

From the skin detection equation [equation 3.2], if the pixels satisfy the equation, then the pixels are considered skin and will be detected in a rectangle window as shown in Figure 3.2.

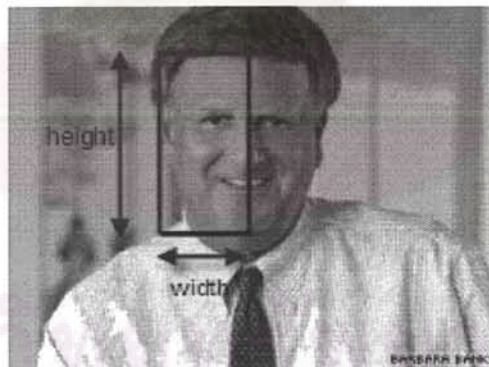


Figure 3.2 : An example of human image using RGB model

Figure 3.2 shows an example after employing the RGB model. The rectangle window represents the location of human face detected from the background of an image. The obtained information are 4 attributes: width, height, size and orientation.

The number of pixels detected inside a rectangle window will be used to compute these 4 attributes: height, width, size and orientation. The height is length of



vertical. The width is length of horizon. The size obtains from the size of skin pixels. Finally, the orientation is computed from height divided by width. If the output value is less than 1, then assign the orientation to 0, otherwise assign to 1. Since a human face has height more than width, result of height divided by width must be more than 1. The example values from the above image are as follows:

#### Information from the low level extraction using RGB model on Figure 3.3

- Height = 108
- Width = 53
- Size = 3087
- Orientation =  $\frac{Height}{Width} = \frac{108}{53} = 2.04 = 1$

In this case, the orientation is 2.04 which is greater than 1. Therefore, the value of 1 will be reassigned as the orientation value.

#### The YCbCr color model

In the YCbCr color space, chrominance components are represented by Cb and Cr values. Thus, skin color model can be derived from these values. For this study, the following threshold [18] is applied.

$$Map_{cb\&Cr(x,y)} = \begin{cases} 255, & Cb \in R_{Cb} \cap Cr \in R_{Cr} \\ 0, & otherwise \end{cases} \quad [Equation 3.4]$$

where  $x = 1, 2, \dots, M/2$  and  $y = 1, 2, \dots, N/2$ .

The equation 3.4 shows skin color pixels identified by the presence of a certain set of Cb and Cr values which corresponding to the respective ranges of  $R_{Cb}$  and  $R_{Cr}$  values only. The suitable value of cb component is in the range [77 127], whereas, cr is in the range [133 173] of skin color. Otherwise, the pixel is classified as non skin color. The result for low level extraction using YCbCr color space are shown in Figure 3.3.



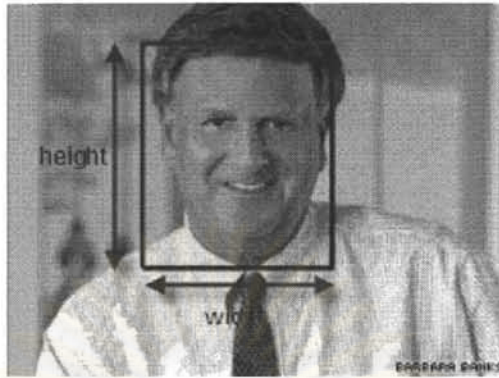


Figure 3.3 : An example of human image using YCbCr model

Figure 3.3 shows an example after employing the YCbCr model. The rectangle window represents the location of human face that the program detected from the background. The obtained information consists of 4 attributes: width, height, size and orientation (same attributes as the RGB model). Using the same image but different skin model, the result is shown as in Figure 3.5. The values of 4 attributes computed using YCbCr model are as follows:

#### Information from the low level extraction using YCbCr model

- Height = 131
- Width = 111
- Size = 8959
- Orientation =  $\frac{Height}{Width} = \frac{131}{111} = 1.18 = 1$

In this case, the orientation is 1.18 which is greater than 1. Therefore, the value of 1 is reassigned as the orientation value.

#### 3.3.3 High level data extraction

The high level data extraction uses the input from the text surrounding images from section 3.3.1. Such text can be further refined through name entity detection to identify a person. To accomplish this high level data extraction, the system collects words from tags, then words are analyzed using lexitron [19]. Lexitron is a Thai-English dictionary which is developed by Nectec.

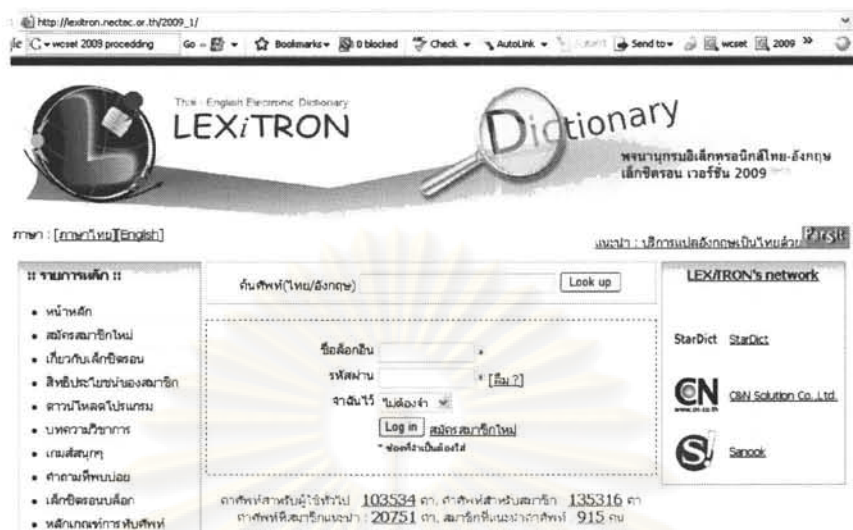


Figure 3.4 : Website dictionary corpus.

Figure 3.6 shows the Website dictionary corpus. The vocabularies from the dictionary are composed of 83,223 words, which are used to analyse words from tags surrounding images. The associated information and tags employed in the experiment are:

1. File name :

It uses to define an image file name. The file name allows one to know additional information about the image. The file name can be a useful way to understand the thinking associate to an image or what the image is about. For example, if an image has the file name named "car.jpg", then it might help to guess what is that image.

2. Caption name :

The caption tag allows users to know the description of an image of the web page. Generally, the caption is appeared between <CAPTION> and </CAPTION> tag. In summary, the caption name is the field that contains text and typically description accompany with an illustration of an image in a very specific way. It shows some degree of control images. Thus, some applications (or people) find that it is convenient to read only the caption. Since it contains the repeated information about the image.

3. Title name :

The title tag shows between the <TITLE> and </TITLE> when viewing the web page. The title tag is used as an alternative text for an image. It is an author-defined text for anyone who visits the page with a browser that cannot display images. The title name is worth focusing in relation to the image that shows in the page.

#### 4. Role and position :

Define role and position such as Secretary, Commentary, Prime minister, CEO. These words, typically, start with capital letters and can be acquired from title and caption tags.

In this thesis, the system also considers words with the capital letters. Since there is a better chance that it might indicate the person name, role or position. However, this high level information will be integrated with the low level information to classify an image of human. The following is an example of extracting high level information from an image.

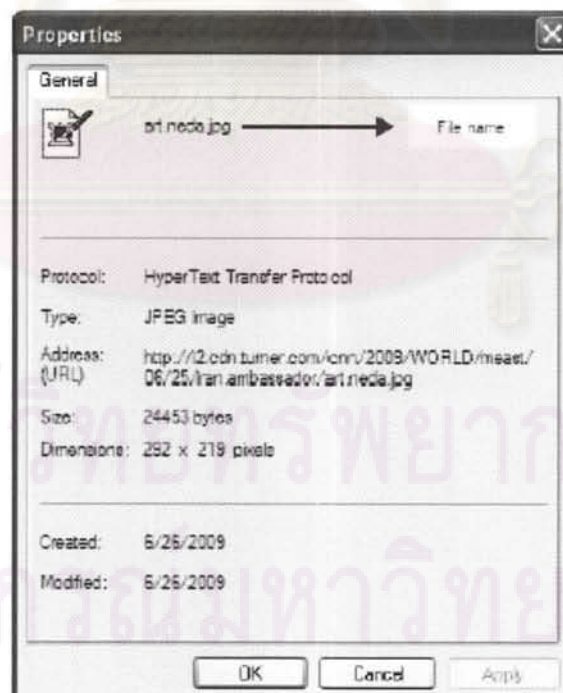


Figure 3.5 : An example of file name



Figure 3.5 shows an example of an image file name. Here, the file name is kept in the system as neda.jpg. Then, the word 'neda' is compared with the database corpus of the dictionary for the proper name. If the compared word is a proper name, set the word value to 1, Otherwise, the value is set to 0.

To simplified the process of wording on the caption name (description or title below an image illustration) which may contain many words. Here, words with capital letters will be examined. Figure 3.6, for example, the caption with capital letters consists of "Neda, Agha-Soltan, Tehran, Saturday". Each word in this caption will be analyzed. Start with "Neda", the word will be checked in the database corpus whether it is the proper name. If it is a proper name, set the word value to 1. Otherwise, the word value is set to 0. In this case, "Nada" is a proper name so the word value is set to 1. Next, "Agha-Soltan" is verified with the corpus for the proper name. If this word is the proper name, set the word value to 1. This process continue for every word in the caption. The final word value is the result from the OR operation of all the proper name word values. This means that if one of the words is the proper name. The final word value is set to 1. Otherwise, the final word value is set to 0.



Neda Agha-Soltan, 26, was shot to death in Tehran on Saturday  
Caption name

Figure 3.6 : An example of the caption name

The title name is processed in the similar manner as the caption. Since the title name may contain many words, to simplify the process only words with capital



letters will be evaluated. Figure 3.7, displays an example of title name tag which can be read from the source code in the HTML file.

**<title>Iranian envoy: CIA involved in Neda's shooting? - CNN.com</title>**

Figure 3.7 : An example of title name

To check the proper name for words appeared in title tag, words: Iranian, CIA and Neda will be evaluated using lexitron [19]; a dictionary corpus developed by Nectec. Each word is verified one by one for the proper name. The word value is set to 1 if it is the proper name. Otherwise, the word value is set to 0. Again, the final word value is the result from the OR operation of all the word values in the title name. This means that if there is a least one proper name, the word value is set to 1. Otherwise, the word value is set to 0.

To process role and position, words obtained from both caption and title tags are investigated. From the above example, there are 6 words (2 words from the title tag and, and 4 words from caption) are investigated. Note that, the role and position words are the list of vocabularies that represent role and position. Then, each word is analyzed with the dictionary program. Examples of role and position words are Secretary, Commentary, Prime Minister, and Singer.

Table 3.1 illustrates an example of the high level data extraction. There are 4 different attribute names; filename, caption, alt and role, position. The value can be either 1 or 0. When 1 denotes that there is a proper name in each attribute type. Otherwise, the value is 0.

Table 3.1 : A sample of high level extraction.

Attributes name	Value
Image filename field	1
Image caption name field	1
Image Alt name field	1
Image role and position name field	0

In the end, there will be 4 attribute values from this high level data extraction. When combined with other 4 attribute values from the low level data extraction. There will be altogether 8 attribute values.

These 8 attributes will be employed as input for training the neural network to classify a single human face from other images.

### 3.3.4 Stage two: Training the network

Neural networks have been applied in many pattern classification problems such as numerical hand written recognition. The neural network with back-propagation training algorithm is shown in Figure 3.11. From the figure, the network consists of 3 layers: input layer, hidden layer and output layer. The input layer contains 8 input nodes, the hidden layer has 4 nodes and the output layer has 2 nodes. In the training stage, the output classes are face and non face which are known in advance (this is termed supervised training). The output node is assigned the value of 1 for face and the value of 0 for non face human image.

#### 3.3.4.1 Input Data

In the experiment, there are 400 color images of face and non face human images. Human images consist of various skin tone from many races such as European, Asian, and African. The non face images include landscapes, animals, and things. The data sets are partitioned into 2 parts, 60% for the training set and 40% for the testing set. Thus, the training images are composed of 240 images. Therefore, there are 160 images left for testing the performance of the network. Each image consists of 8 attribute values from the low and high level data extraction. These attribute values will be used as the input patterns for the network.

#### 3.3.4.2 Output Data

The neural network learns to classify the output of face and non face human image. A face is represented by the output node values "1" and "0" for the non face. The experimental simulations are run on the well-known machine learning suite: WEKA [20].

Figure 3.11 shows a typical architecture of a multilayer perceptron network. A Multi-layers Perception (MLP) is a particular kind of artificial neural network. The MLP is used extensively to solve a number of different problems, including pattern recognition and interpolation. Each layer is composed of neurons, which are interconnected with each other by weights. In each neuron, a specific mathematical function called activation function accepts input from previous layer and generates output for the next layer. In the experiment, utilized activation function is the Hyperbolic tangent sigmoid transfer function [21]. The MLP is trained using a standard back-propagation algorithm.

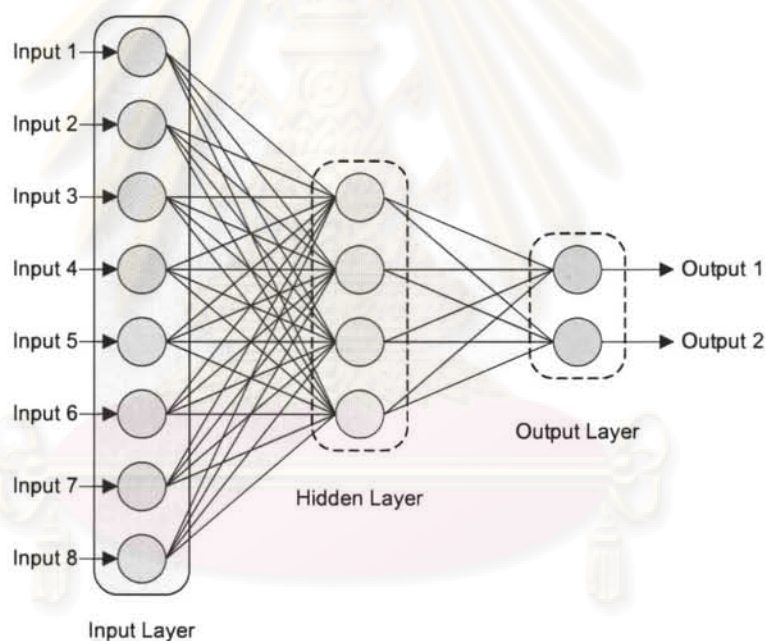


Figure 3.8 : A typical Multilayer Perceptron ANNs Architecture.

Choosing the number of hidden units is an important factor. There are several publications discussed the number of hidden units, such as Elisseeff, et al [22]. The following equation is used as a guideline to estimate the number of hidden unit (H) :

$$H \geq \frac{n - m}{m(k + 2)}$$

K is the number of inputs equal to the number of attributes. N is the sample size (400). In order to light over fit the data, there must be fewer than m cases for each parameter.

Normally,  $m$  set to 10. The value of hidden units can be calculated by replacing each parameter with its corresponding value using the above formula as following:

$$H \geq \frac{400 - 10}{10(8 + 2)} \approx 4$$

The number of hidden units initially are 4. However, hidden units are varied to get the optimal results. The architecture of the network in this study finally becomes 8 input nodes, 8 hidden nodes and 2 output nodes (8:8:2). The learning rate and momentum are set to 0.2 and 0.3, respectively.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## CHAPTER IV

### EXPERIMENTAL AND RESULTS

#### 4.1 Experimental results

The experimental simulations are run WEKA 3.6. Two studies are conducted: 1) Compare between two different color models: RGB and YCbCr. Which color model is more suitable for detecting human skin tone.

2) Compare the performance of the classification with and without text information (high level data extraction) using the better color model from the first study. (Here, the result shows that YCbCr model provides the better result than the RGB model)

##### 4.1.1 Experiment on Images using two color models: RGB and YCbCr

As mentioned earlier, the data extracted from images (low level data extraction) obtain from skin detection. The skin detection can be performed using various color models such as RGB, YCbCr, and HSV etc. However, the central focus of this research is on only RGB and YCbCr model. To find out which color model is more efficient in detecting human skin tone. The results of the facial classification using RGB and YCbCr model are shown in Table 4.1 and Table 4.2 respectively.

Table 4.1 : Human facial image classification results using RGB.

Prediction outcome	Actual	
	Facial image	Non facial image
Facial image	73	15
Non facial image	0	72

Table 4.1 shows the confusion matrix of the network performance on the RGB color model. From the table, there are 73 human images and the network classifies correctly as the human images. In addition, there are 72 non-facial human images, in which the network also classifies correctly as the non-facial images. The accuracy performance of the system with the RGB model is calculated by the following formula:

$$ACC = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}}$$

$$ACC = \frac{73 + 72}{73 + 15 + 72 + 0} = 90.625\%$$

This means that with RGB model, the accuracy percentage for classifying the human image is 90.625%

Table 4.2 : The face classification result using YCbCr.

Prediction outcome	Actual	
	Facial image	Non facial image
Facial image	100	2
Non facial image	6	52

Table 4.2 shows the confusion matrix of the network performance on the YCbCr color model. From the table, there are 100 human images and the network classifies correctly as the human images. In addition, there are 52 non-facial human images, in which the network also classifies correctly as the non-facial images. The accuracy performance of the system with the YCbCr model is calculated using the following formula:

$$ACC = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}}$$

$$ACC = \frac{100 + 52}{100 + 2 + 52 + 6} = 95\%$$

The accuracy performances from Table 4.1 and Table 4.2, are 90.625% and 95% respectively. It can be easily seen that the YCbCr model has the better performance than RGB model in detecting human skin tone.

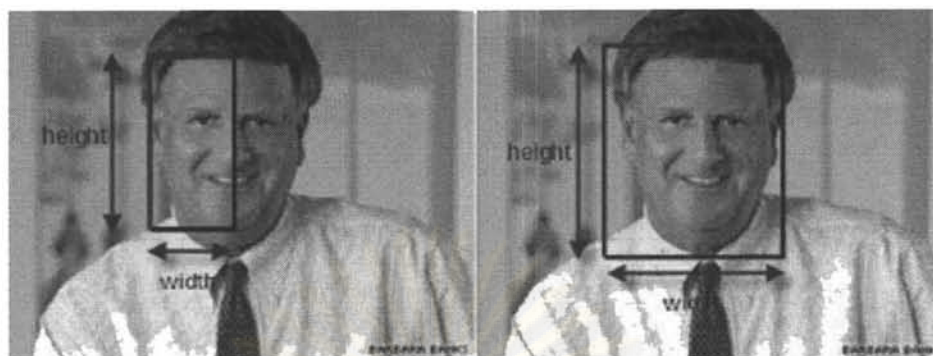


Figure 4.1 : An example of detecting human skin tone using RGB and YCbCr model.

Figure 4.1 shows the comparison between RGB and YCbCr skin color detected by the system. On the left hand side figure, the RGB color space is found to represent only partial human face. This might be because the detected skin region of RGB is discontinuous due to lighting effects so this leads to the missed skin pixels. In addition, RGB color space contains three normalized components  $r$ ,  $g$  and  $b$ , which is the pure color without information about the luminance. On the other hand, YCbCr color model has information about the luminance [23] by using  $Cb$  and  $Cr$  to distinct color ranges for skin region. Therefore, the YCbCr shows a better performance in detecting human skin tone than the RGB model.

#### 4.1.2 Experiment with and without text surrounding images

To verify that the text surrounding images are useful information for classifying the human image, this study compares the accuracy percentage with text and without text information (information from high level data extraction).

In this experiment, the input data consists of 400 color images and these images are automatically collected from the CNN website. The images are further compress in the format of GIF files. The ratio of the training set and testing set is 60:40. Therefore, 400 images are divided into 240 images for training and 160 images for testing the performance of the network.

Table 4.3 : The face classification result using YCbCr with text information (high level description).

Prediction outcome	Actual	
	Facial image	Non facial image
Facial image	100	2
Non facial image	6	52

Table 4.4 : The face classification result using YCbCr without text information (high level description).

Prediction outcome	Actual	
	Facial image	Non facial image
Facial image	78	6
Non facial image	9	67

Table 4.3 and table 4.4 shows the confusion matrix of the predictive results of the network with text information (high level description) and without text information.

The accuracy percentage of the system performance with text information from Table 4.3 is calculated by the following formula:

$$ACC = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}}$$

$$ACC = \frac{100 + 52}{100 + 2 + 52 + 6} = 95\%$$

The accuracy percentage of the system performance without text information from table 4.4 is calculated by the following formula:

$$ACC = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}}$$

$$ACC = \frac{78 + 67}{78 + 6 + 67 + 9} = 90.63\%$$



The results of the accuracy percentage from Table 4.3 and Table 4.4 are 95% and 90.63%, respectively. Thus, it can be seen that YCbCr with text information surrounding image has better performance than without text information. This indicate that text surrounding images provide useful information for classifying human images.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## CHAPTER V

### CONCLUSION AND FUTURE WORKS

#### 5.1 Conclusion

In this study, the predictive model is presented to classify human facial images. The model is performed on both the image data (low level description) and text surrounding images (high level description). The image information (height, width, size and orientation) and the text surrounding image information (filename, title, caption, and position) are automatically obtained by an author's developed program or crawler. The results illustrate that YCbCr model with 95% accuracy has a better performance than RGB model with 90.625% accuracy in detecting the human skin tone. In addition, when compare the performance of the YCbCr model with text information, the accuracy is 95% accuracy. Whereas, without text information the accuracy is only 90.63%. This illustrate that the text information can enhance the performance of the system.

It is concluded that metadata can add value to image classification. The summarized detail of the experiment is shown in Figure 5.1.

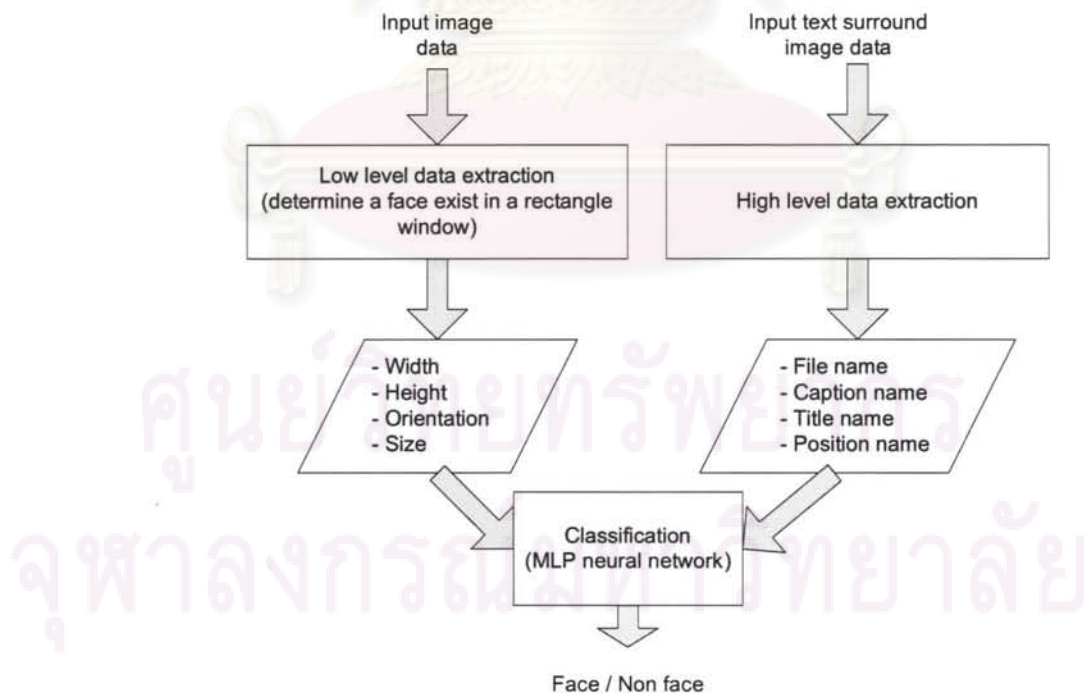


Figure 5.1 : The experimental process.

## 5.2 Discussions

The YCbCr has a better performance than the RGB model. Appendix A illustrates more comparisons on the window/ frame of the detected skin-tone between these two models. The RGB color space is found to represent only partial human face. This might be because the detected skin region of RGB is discontinuous due to lighting effects so this leads to missing skin pixels. In addition, RGB color space contains three normalized components r, g and b, which is the pure color without information about the luminance. On the other hand, YCbCr color model has information about the luminance. In addition, the YCbCr model utilizes Cb and Cr to distinct color ranges for skin region. Therefore, the YCbCr shows a better performance in detecting human skin tone than the RGB model. Note that, the skin tone of human images employed in the experiments containing many races such as European, Asian and even African.

The study proposed a new simple and efficient technique in classifying a human facial image using the basic information of the image itself and the metadata surrounding the image. The experimental results indicated that metadata can enhance the performance of the classifications. The outcome from the experiments can enhance the understanding on how images should be indexed to facilitate the retrieving rate on the human facial images. In addition, this can assist the image resource management for online searching system.

## 5.3 Future works

There are several interesting issues need be addressed such as:

1. Changing the YCbCr color model, which is employed in the low-level data extraction, to a different color model may improve the performance of the system.
2. A more controlled vocabulary for describing specific human facial images may improve the ability to discover image resources

Finally, there should be a balance between what users ask for and what the metadata can support. These developments have heightened the need for effective

image retrieval techniques. If the metadata are consistent, it will allow the search engine to generate good hit lists.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## REFERENCES

- [1] Arlene G. Taylor. **The organization of information**. Chicago: American Library Association. 2004.
- [2] M.H. Yang, D.J. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence** (2001): 34 - 58.
- [3] S. G. Tzafestas. Advance in intelligent system: concepts, tools and applications. **Proceedings of international series on microprocessor-based and intelligent systems engineering** (1999): 288 - 300.
- [4] Jitendra Malik, Serge Belongie, Thomas Leung and Jianbo Shi. Contour and Texture Analysis for Image Segmentation. **Proceeding of International Journal of Computer Vision** (2001): 7 – 27.
- [5] S.G. Tzafestas. Advance in intelligent systems: concepts, tools and applications. **Proceeding of International series on microprocessor-based and intelligent systems engineering** (1999): 288 – 300.
- [6] Stephen J. Sangwine and Robin E.N. Horne. **The color image processing handbook**. The United kingdom: Chapman & Hall. 1998.
- [7] S.P. Khandait, R.C. Thool. Face Processing using Skin Detection Algorithm. **International Journal of Computational Intelligence and Healthcare Informatics** (2009): 77- 88.
- [8] R-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain. Face Detection in Color Images. (PAMI) **IEEE Transactions on Pattern Analysis and Machine Intelligence** (2001): 696 – 706.
- [9] S. Kannumuri and A.N. Rajagopalan. Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information. (ICVGIP) **Indian Conference on Computer Vision, Graphics and Image Processing** (2002): 312 – 317.
- [9] Rowley H., Baluja S. and Kanade T. Neural Network-Based Face Detection. **Proceeding of IEEE Conference on Computer Vision and Pattern Recognition** (1996): 203 – 207.

- [10] McCulloch & Pitts. A Logical Calculus and the Ideas Immanent in the Nervous Activity. **Bulletin of Mathematical Biophysics** (1943): 115 – 133.
- [11] Rowley H., Baluja S. and Kanade T. Neural Network-Based Face Detection. **Proceeding of IEEE Conference on Computer Vision and Pattern Recognition** (1996): 203 – 207.
- [12] Y. H. Nam, Y. Y. KIM, H. T. Kim. Automatic Detection of Nausea Using Bio-Signals During Immerging in A Virtual Reality Environment. **International Conference of the IEEE Engineering in Medicine and Biology Society** (2001): 2013 – 2015.
- [13] Diego Andina, Duc Truong Pham. **Computational intelligence for engineering and manufacturing**. Netherland: Springer, 2007.
- [14] Luis A, Trejo and Carlos Sandoval. Improving Back-Propagation: Epsilon-Back-Propagation. **Proceeding of an International Workshop on Artificial Neural Networks** (1995): 427 – 430.
- [15] HyunKyung Park, ByeongCheol Yoo, JeGoon Ryu and Toshihiro Nishimura. The Speckle Reduced Ultrasound Images Using Cellular Neural Network with Effective Detection of Active Contour Model. **Proceeding of the World Congr on Engineering and Computer Science** (2008): 187 – 192.
- [16] Kazuhiro Kohara. Selective-Learning-Rate Approach for Stock Market Prediction by Simple Recurrent Neural Networks. **Knowledge-based intelligent information and engineering systems: 7<sup>th</sup> International Conference** (2003). 140 – 147.
- [17] Choochaiwattana W., Nirantlumpong W. and Spring M.B. Web image classification algorithm: A heuristic rule-based approach. **The Second International Conference on Internet Technologies and Application** (2007): 202 – 224.
- [18] D. Chai and K. N. Ngan. Face segmentation using skin-color map in videophone application. **IEEE Transactions on Circuits and System for video Technology** (1999): 551 – 564.
- [19] Lexitron [online]. Available from: [http://lexitron.nectec.or.th/2009\\_1](http://lexitron.nectec.or.th/2009_1) [2009, March 1].

- [20] Weka [online]. Available from: <http://www.cs.waikato.ac.nz/~ml/weka/index.html> [2005, June 1].
- [21] A. Ngaopitakkul and A. Kunakorn. Selection of Proper Activation Functions in Back-propagation neural networks algorithm for Transformer Internal Fault Location. (IJCNS) *International Journal of Computer and Network Security* (2009): 690 – 699.
- [22] Matignon, R. *Neural Network Modeling*. EAuthorhouse. 2005.
- [23] Cai J., Goshtasby A., and Yum. Detecting Human Faces in Color Images. *Proceedings of International Workshop on Multi-Media Database Management Systems* (1998): 124 – 131.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



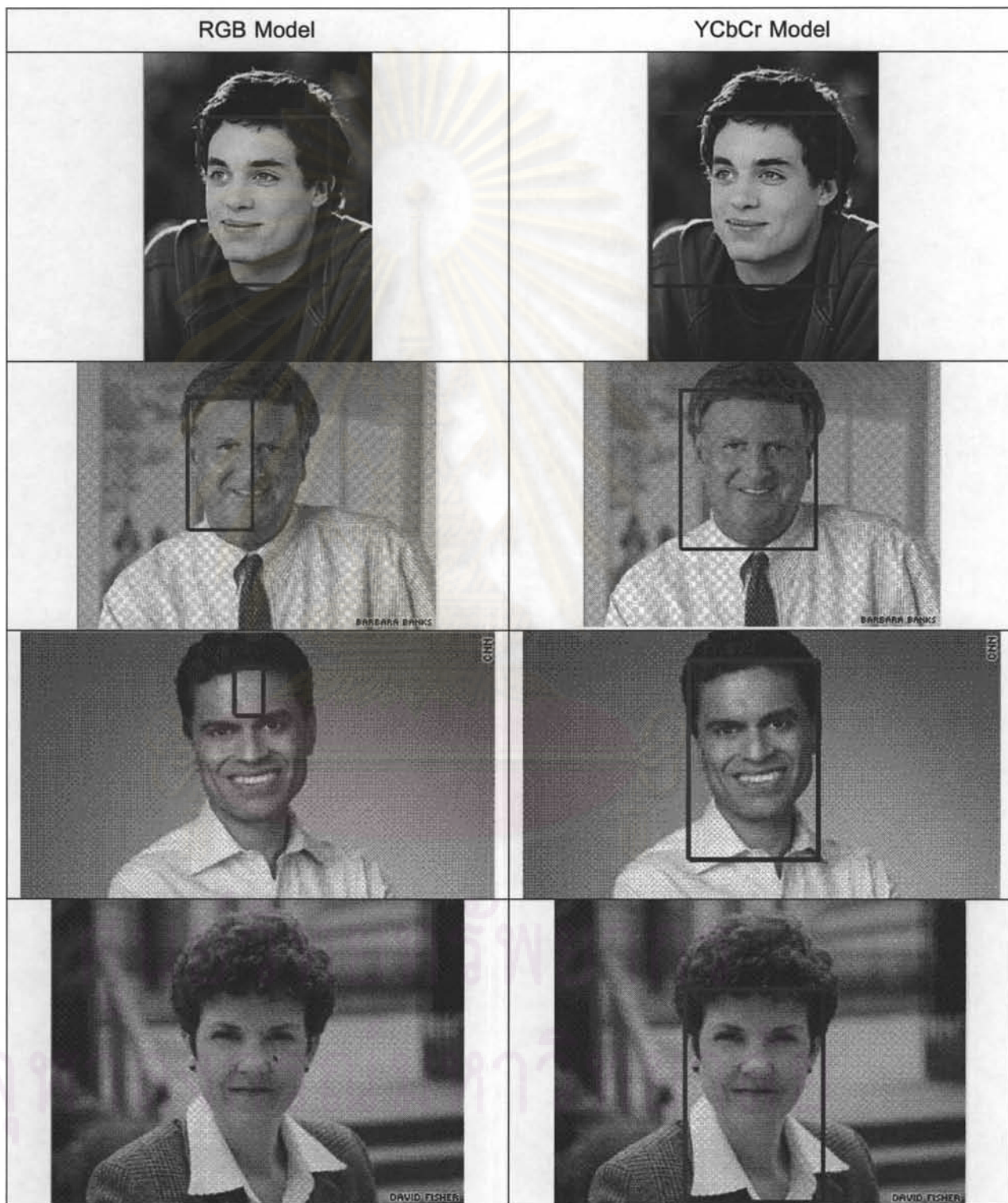
APPENDIC

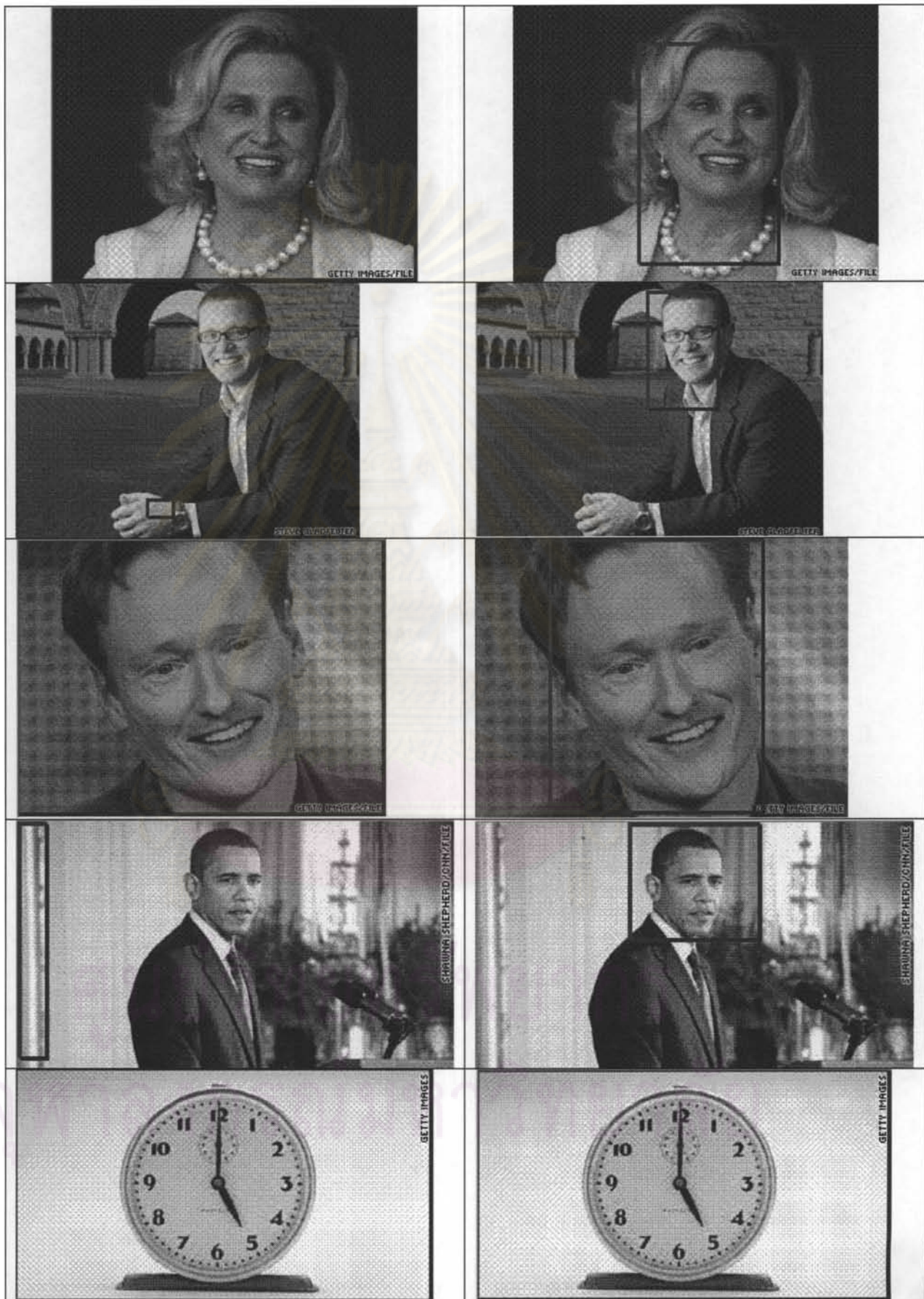
ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## Appendix A

Comparison between RGB and YCbCr model on detecting human facial image.





The output obtained in the same image from the low level extraction as the example in above images. The frame shown the face area detected from background and compare between the RGB and YCbCr model. Some window of faces detected using RGB are missed so the model using YCbCr detected more correctly than RGB model.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



## VITA

Chutimon Thitipornvanid was born in 1982. She received a Bachelor Degree in Science (Majoring Computer Science) from Prince of Songkla University in 2005. She is working as Quality assurance for DST international, Bangkok and is also pursuing a Master degree in computer science.

### Publication

Chutimon, T., and Siripun, S., "Prediction of a Human Facial Image by ANN using Image Data and Content on Web Pages" Proceedings of 2009 International Conference on Computer Science and Software Engineering (WASET), August 26<sup>th</sup>-28<sup>th</sup>, 2009

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย