

การระบุนชนิดและท่าของภาพกีฬาที่มีพื้นหลังซับซ้อนโดยใช้ตัวอธิบายฟูเรียร์



นางสาวปิยะนันท์ พนากานต์

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต


สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

IDENTIFYING TYPES AND GESTURES OF SPORTS IMAGES WITH COMPLEX  
BACKGROUND BY FOURIER DESCRIPTOR



Miss.Piyanan Panakarn

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy Program in Computer Science  
Department of Mathematics

Faculty of Science

Chulalongkorn University


Academic year 2010

Copyright of Chulalongkorn University


Thesis Title IDENTIFYING TYPES AND GESTURES OF SPORTS IMAGES  
WITH COMPLEX BACKGROUND BY FOURIER DESCRIPTOR  
By Miss Piyanan Panakarn  
Field of Study Computer Science  
Thesis Advisor Professor Chidchanok Lursinsap, Ph.D.  
Thesis Co-advisor Suphakant Phimoltares, Ph.D.

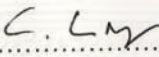
---


Accepted by the Faculty of Science, Chulalongkorn University in Partial  
Fulfillment of the Requirements for the Doctoral Degree

  
..... Dean of the Faculty of Science  
(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

  
..... Chairman  
(Associate Professor Peraphon Sophatsathit, Ph.D.)

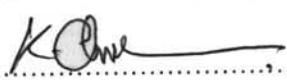
  
..... Thesis Advisor  
(Professor Chidchanok Lursinsap, Ph.D.)

  
..... Thesis Co-advisor  
(Suphakant Phimoltares, Ph.D.)

  
..... Examiner  
(Professor Boonserm Kijirikul, Ph.D.)

  
..... Examiner  
(Assistant Professor Saranya Maneeroj, Ph.D.)

  
..... External Examiner  
(Associate Professor Kosin Chamnongthai, Ph.D.)

  
..... External Examiner  
(Assistant Professor Krisana Chinnasarn, Ph.D.)

ปิยะนันท์ พนากานต์ : การระบุชนิดและท่าของภาพกีฬาที่มีพื้นหลังซับซ้อนโดยใช้ตัว  
อธิบายฟูเรียร์ (IDENTIFYING TYPES AND GESTURES OF SPORTS IMAGES  
WITH COMPLEX BACKGROUND BY FOURIER DESCRIPTOR) อ. ที่ปรึกษา  
วิทยานิพนธ์หลัก: ศาสตราจารย์ ดร. ชิดชนก เหลือสินทรัพย์, อ.ที่ปรึกษาวิทยานิพนธ์  
ร่วม: อ.ดร. สุภกานต์ พิมลธเรศ, 83 หน้า.

การจำแนกภาพด้วยเนื้อหาความหมายเป็นปัญหาที่น่าท้าทายในปัจจุบัน เทคนิคหลาย  
ชนิดเช่น การจำแนกภาพด้วยเนื้อหาของภาพ การจำแนกภาพด้วยข้อความอธิบายรูปภาพ และ  
วิธีการจำแนกภาพที่ใช้ทั้งเนื้อหาของภาพและคำอธิบายของภาพได้พัฒนาขึ้นเพื่อแก้ปัญหานี้  
เทคนิคหลายชนิดได้แก่ การลบพื้นหลังภาพ การตรวจจับบริเวณที่น่าสนใจและการจดจำวัตถุ  
ได้นำมาใช้ร่วมกันเพื่อนำไปสู่การเพิ่มประสิทธิภาพในการจำแนกภาพอย่างมีความหมาย โดย  
การรวมเอาหลายเทคนิควิธีมารวมกัน จากความสามารถในการประยุกต์ของเทคนิคผสมเหล่านี้  
งานวิจัยนี้ได้แก้ปัญหการจำแนกภาพด้วยความหมายโดยเน้นในส่วนของภาพกีฬา วิธีการที่  
นำเสนอที่ใช้ท่าทางของผู้เล่นเป็นพื้นฐานในการระบุเพื่อจำแนกภาพรับเข้าเพื่อแสดงถึงท่าทาง  
ของผู้เล่น ตัวอย่างเช่น นักเบสบอลมีท่าทางการเล่นที่เรียกว่า 'ท่าตี' และ 'ท่าโยน' ที่คล้ายคลึง  
กัน วิธีการที่นำเสนอสำหรับการแก้ปัญหานี้ประกอบด้วย วิธีการกำจัดพื้นหลังที่มีความ  
ซับซ้อนออกแบบอัตโนมัติและวิธีการจำแนกท่าทางของผู้เล่นบนข้อมูลกีฬาประเภทเพื่อวัด  
ความสามารถในการประยุกต์ของวิธีการ วิธีการที่นำเสนอมีความคงทนต่อภาพที่มีการเปลี่ยน  
แปลงขนาดและมีการหมุนของผู้เล่น

ภาควิชา ..... ภูมิศาสตร์ ..... ลายมือชื่อนิสิต..... มิเชลล์ พนากานต์  
สาขาวิชา ..... วิทยาการคอมพิวเตอร์ ..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก..... C. L. S  
ปีการศึกษา ..... 2553 ..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม ..... Split Print



## 4973830523 : MAJOR COMPUTER SCIENCE

KEYWORDS : SPORT IMAGE CLASSIFICATION / POSTURE CLASSIFICATION / SEMANTIC SPORT IMAGE CLASSIFICATION / IMAGE CLASSIFICATION / NEURAL NETWORKS / BACKGROUND ELIMINATION

PIYANAN PANAKARN : IDENTIFYING TYPES AND GESTURES OF SPORTS IMAGES WITH COMPLEX BACKGROUND BY FOURIER DESCRIPTOR, ADVISOR: PROFESSOR. CHIDCHANOK LURSINSAP, Ph.D., CO-ADVISOR : SUPHAKANT PHIMOLTARES, Ph.D., 83 pp.

Image classification with semantic contents is a challenging problem in recent years. Various techniques such as content based image classification, text based image classification, and combined content and text based image classification have been developed to solve this problem. To boost a high of semantic image classification performance, several techniques such as background elimination, region of interest detection, and object recognition have been incorporated. In an attempt to explore the applicability of such combined techniques, this research undertakes the semantic image classification problem by focusing on sport image domain. The proposed technique use player's posture as the identification basis to classify the input image is the posture of player classification. For example, a baseball player image should show the 'hitting' and 'throwing' postures. To overcome this problem, the proposed method utilizes automatic complex background elimination and posture classification based on six different sports to gauge the applicability of the approach. The proposed method is robust to scaling and rotation player image.

Department : Mathematics

Field of Study : Computer Science

Academic Year : 2010

Student's Signature

Advisor's Signature

Co-advisor's Signature

*Piyanan Panakarn*

*C. L. L.*

*Suphakant Phimoltares*

## Acknowledgements

I am very appreciates to my advisors: Professor Dr. Chidchanok Lursinsap and Dr. Suphakanth Phimoltares. Both of them introduce me in the way of doing research and pointing to ordering the priority of tasks. Moreover, they introduce me to the scholarship for last two years of my study which is very important to me.

I would like to thank Associate Professor Peraphon Sophatsathit for his discussions and suggestions to my research solving.

I would like to thank AVIC Lab and all members for helping all discussions and conveniences of my research study. Furthermore, I would like to thank Associate Professor Suchada Siripant to her providing of facility and suggestions for any problems during my studying.

For the main financial support, I would like to thank the supporting of the Office of Higher Education Commission, Ministry of Education, Thailand through Faculty of Science, Ubon Ratchathani University. I would like to thank the supporting of the Center of Excellent in Mathematics, CHE, Sri Ayutthaya Rd., Bangkok, 10400, Thailand. In addition, I would like to appreciate for the scholarship of Science and Technology Innovation Support Grant from Chulalongkorn University.

Lastly, I would like to thank my parents and family for their encouragements.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

# Contents

	Page
<b>Abstract (English)</b> .....	<b>v</b>
<b>Acknowledgments</b> .....	<b>vi</b>
<b>Contents</b> .....	<b>vii</b>
<b>List of Tables</b> .....	<b>ix</b>
<b>List of Figures</b> .....	<b>x</b>
Chapter	
<b>I INTRODUCTION</b> .....	<b>1</b>
1.1 Objectives .....	1
1.2 Scope and Limitation .....	2
1.3 Contributions .....	2
1.4 Research Methodology .....	2
1.5 Dissertation Organization .....	3
<b>II RELATED WORK</b> .....	<b>4</b>
2.1 Sport image classification .....	4
2.1.1 Combination of textual information and image content .....	4
2.1.2 Geometrical information .....	5
2.1.3 Content-based image retrieval .....	5
2.1.4 Adaptive of fusion neural network .....	5
2.2 Semantic image classification .....	7
2.2.1 Semantic situation in sport game .....	7
2.2.2 Classifying events by scene and object recognition .....	7
2.2.3 Combined method .....	8
2.3 Tools reviews .....	8
2.3.1 YCbCR Color Model .....	8
2.3.2 HSV Color Model .....	9
2.3.3 Discrete Cosine Transform (DCT) technique .....	10
2.3.4 Discrete Wavelet Transform (DWT) technique .....	11
2.3.5 2D Discrete Fourier Transform (2D-DFT) .....	11
2.3.6 Singular Value Decomposition (SVD) .....	12
<b>III SPORTS IMAGE CLASSIFICATION</b> .....	<b>14</b>
3.1 Majority Color Extraction .....	14
3.2 Descriptor for Complex Background .....	15
3.3 Posture Descriptor .....	16
3.4 SVD on image representation concept .....	18
3.5 Algorithm of sport image classification .....	19
3.5.1 Examples of image features .....	22

Chapter	Page
<b>IV POSTURE CLASSIFICATION .....</b>	<b>26</b>
4.1 Blurred background elimination .....	26
4.2 Proposed irrelevant area elimination .....	29
4.2.1 Proposed player area selection .....	32
4.2.2 Proposed semantic feature extraction .....	34
<b>V EXPERIMENTAL RESULTS .....</b>	<b>44</b>
5.1 The overall system .....	44
5.2 Sport image classification .....	44
5.2.1 Discussion on Features Comparison on NNC, FNN, and RBF .....	46
5.2.2 Discussion on Distribution of Misclassification .....	46
5.3 Posture classification .....	47
5.3.1 Experimental Set-up for posture classification .....	50
5.3.2 Feature comparison .....	50
5.3.2.1 EH & RS features (EH & RS) .....	50
5.3.2.2 Proposed features (Pro.Feat.) .....	50
5.3.3 Parameter Set-up .....	51
5.3.4 Analysis of sport image classification .....	52
5.3.5 Analysis of posture recognition .....	52
<b>VI CONCLUSION .....</b>	<b>61</b>
6.1 <i>Sport classification</i> .....	61
6.2 <i>Posture classification</i> .....	61
<b>REFERENCES .....</b>	<b>62</b>
<b>Appendix .....</b>	<b>67</b>
APPENDIX A .....	68
<b>Biography .....</b>	<b>72</b>



## List of Tables

Table	Page
3.1 An example of majority colors calculation for field, track, ice, and water backgrounds. . .	15
5.1 The comparison of classification accuracy between EH & RS and our features on three kinds of neural network. . . . .	47
5.2 The classification accuracy in each sport compared between NNC and FNN neural network on our features, 1,000 and 5,000 epochs. . . . .	48
5.3 The classification accuracy in each sport compared between NNC and FNN neural network on our features, 10,000 and 20,000 epochs. . . . .	49
5.4 Postures definition. . . . .	51
5.5 The percentage of accuracies in image classification and misclassification of six sports with learning rate 0.1, 1,000 and 5,000 epochs (Notice:ep.=’epochs’, and hd=’hidden nodes’) . . . . .	53
5.6 The percentage of accuracies in image classification and misclassification of six sports with learning rate 0.1, 10,000 epochs. (Notice:ep.=’epochs’ and hd=’hidden nodes’) . . . .	54
5.7 The average accuracies of postures classification and misclassification compare between proposed features(Pro.Feat.) and EH & RS features(EH&RS). (Notice, BB=’baseball’, BK=’basketball’, FT=’field & track’,ep.=’epochs’, hd=’hidden nodes’, div.=’dividing number’, and rez.=’resizing number’) . . . . .	55
5.8 The average accuracies of postures classification and misclassification. (Notice, SC=’soccer’, SK=’skiing’, SW=’swimming’,ep.=’epochs’, hd=’hidden nodes’, div.=’dividing number’, and rez.=’resizing number’) . . . . .	56
5.9 The comparison table of average accuracy between proposed features and EH & RS based on 80, 90, and 100 hidden nodes (hd). . . . .	57

## List of Figures

Figure	Page
2.1 The overall architecture of FNNC. . . . .	6
3.1 Examples of transforming RGB color format to YCrCb color format and the comparison of outstanding shapes from Cb and Cr color space. Top row: an example of blue color image denoted as the result of Cb or blue-difference value. Bottom row: an example of Cb and Cr color spaces for black clothes. . . . .	17
3.2 Overall architecture illustrated on the feature vector. . . . .	19
3.3 Performed input vectors from SVD matrices on S-Cb, S-Cr and S-DCT. . . . .	20
3.4 The overall system of sport image classification module. . . . .	20
3.5 Example of 64-bins color histogram feature in six sports. The left column is input image and right column is 64-bins color histogram. Top row: baseball and basketball. Middle row: track & field and soccer. Bottom row: skiing and swimming. . . . .	23
3.6 Example of YCbCr feature in six sports. The left-top sub-image is the input image and the right-top sub-image is Y color mode. The bottom-left is Cb color mode and bottom-right is Cr color mode. In top row illustrate images from baseball and basketball, in middle row presents images from field & track and soccer, and in bottom row shows images from skiing and swimming. . . . .	24
3.7 Example of S-DCT feature in six sports. The left column of sub-image is the input image and the right column of sub-image is S-DCT. In top row illustrate images from baseball and basketball, in middle row presents images from field & track and soccer, and in bottom row shows images from skiing and swimming. . . . .	25
4.1 Blurred background elimination (a) The original image (b) Value mode of HSV color model. (c) The results of 2D-DWT eliminated (d) Thresholding applied (e) Gray scaled of the result of threshold applied. (f) Dilating by ball shape; radius and height are 2 and 4 pixels. (g) Dilating by line shape kernel. (h) Dilating by ball shape; radius and height are 3 and 2 pixels. (i) Biggest connected pixels. (h) Grayscale image. . . . .	37
4.2 High-contrast elimination (top row), and Low-contrast elimination (bottom row). . . . .	38
4.3 Area definition in the highlight area. (a) Original image. (b) Blurred background. (c) Player area. (d) High-contrast area. (e) Low-contrast area. . . . .	38
4.4 The player pixels area has been selected by normal distribution corresponding to pixel histogram in horizontal and vertical direction. . . . .	39
4.5 Defining horizontal and vertical images. . . . .	39
4.6 Overall feature extraction architecture. . . . .	40
4.7 An example of images sliced into two to ten columns. . . . .	41
4.8 An example of images sliced into two to ten rows. . . . .	42
4.9 Example images of hitting, throwing, and situation postures in baseball from three proposed algorithms. . . . .	43
5.1 The overall process of our proposed, sport image classification module and semantic sport image classification module. . . . .	45
5.2 The example of feature extraction example for sport image classification a) Cb channel. b) Cr channel. c) 64-bins color histogram. d) S-DCT coefficient. . . . .	50
5.3 Example of misclassified image of basketball. . . . .	57
5.4 Example of basketball images. . . . .	58
5.5 Example of misclassified images of field & track. . . . .	58

Figure	Page
5.6 Example of field & track images. . . . .	59
5.7 Example of misclassified images of swimming. . . . .	59
5.8 Example of swimming images. . . . .	60
1 Example images of soccer are misclassified in baseball. . . . .	68
2 Example images of basketball are misclassified in baseball. . . . .	68
3 Example image of field & track is misclassified in baseball. . . . .	68
4 Example image of skiing is misclassified in baseball. . . . .	69
5 Example image of baseball is misclassified in basketball. . . . .	69
6 Example image of baseball is misclassified in field & track. . . . .	69
7 Example images of basketball are misclassified in field & track. . . . .	69
8 Example images of soccer are misclassified in field & track. . . . .	70
9 Example image of skiing is misclassified in field & track. . . . .	70
10 Example image of swimming is misclassified in field & track. . . . .	70
11 Example images of baseball are misclassified in soccer. . . . .	70
12 Example image of basketball is misclassified in soccer. . . . .	71
13 Example images of field & track are misclassified in soccer. . . . .	71
14 Example images of baseball are misclassified in skiing. . . . .	71
15 Example images of baseball are misclassified in swimming. . . . .	71

# CHAPTER I

## INTRODUCTION

Searching for image in a large database usually requires a tool to retrieve related images in response to a user's query. A good search tool should return accurate results but simple to use. This is by all means a challenging task to undertake. This study focuses on sport images as a case study to achieve such goal. The underlying principles often utilize image matching which is based on contents based image retrieval technique such as feature selection and extraction. The features involved are color, edge, and shape information. The accuracy of classification relies on the effectiveness of feature selection. The most effective classifier in this area is neural network technique which is suitable for supervised learning and high dimensional classification.

Studies revealed that enhancement of feature selection and extraction employed geometry method [1, 2], the color and edge approach [3, 4], and text based approaches [5, 6]. However the results were not yet suitable for sport image classification with respect to the performance and time complexity.

A number of current studies focused on semantic classification [7–9], where components were represented by object labeling technique [10]. This technique provides good results only in certain sports, but is not suitable for posture classification. Despite development of many video detection techniques [11, 12], they can not be applied directly to still image.

In 2009, Kang et al. [4] proposed the classification of sport image by the fusion neural network classification (FNNC). This method considers two properties of equivalent contribution on MPEG-7 descriptors, namely, edge histogram (EH) and region-based shape (RS). The method can classify the type of sports but is not enough to incorporate the semantic of posture of the players. Moreover, FNNC is structurally more complicated than that of conventional neural network classifier (NNC). To address this weakness, this study aims to determine posture classification based on new feature selection and NNC, thereby improving feature selection in ground sport image classification.

### 1.1 Objectives

1. To develop an efficient algorithm for classifying type of sports from a given image.

2. To develop an efficient algorithm for determining the semantic posture of each sport.

## 1.2 Scope and Limitation

1. Only following sports are considered baseball, basketball, field & track, soccer, skiing and swimming. Each input image is RGB color of the size  $397 \times 594$ .
2. The postures are considered in each sport as follows:
  - (a) Baseball : Hitting, Throwing, and Home plate's situation (Situation 1).
  - (b) Basketball : Running, Shooting, and Under goal.
  - (c) Field & Track : Broad Jump, High Jump, Group Running, Track Running, and Throwing.
  - (d) Soccer : Jumping, Kicking One, and Running Two.
  - (e) Skiing : Jumping and Turning.
  - (f) Swimming : Freestyle, Breaststroke, and Butterfly.

## 1.3 Contributions

The proposed posture classification algorithms are reach to valuable of image processing following.

1. Focus area detection applied by wavelet denoising and its inverse.
2. Foreground object detection applied high-contrast and low-contrast area elimination by zeros means concept.
3. Human posture classification applied by fourier transform based on resizing and slicing image.
4. Posture classification is robustness on differences size player size applied by player area detection algorithm.

## 1.4 Research Methodology

In this dissertation, a new framework for semantic posture and type classification in sport image based on automatic of complex background elimination is proposed. Each constituent algorithm will be carried out as follows.

*Classifying sports type:*



1. Review and study the related research in sport image classification and sport video detection.
2. Investigate new features suitably for sport image classification avoiding segmenting of input image.
3. Prepare data and split into training and testing sets to undergo 4-fold cross-validation.
4. Extract features for sport image classification.
5. Train and test dataset with the proposed features.
6. Compare the proposed method of sport image classification with other algorithms.

*Classifying posture of player:*

1. Review and study the related research in semantic classification of sport image and sport video.
2. Develop algorithm to remove irrelevant background.
3. Develop algorithm to select player area.
4. Investigate the suitable features and develop algorithm for feature extraction.
5. Extract features from the result image obtained from method of classifying sports.
6. Train and test the data sets with the proposed features.

## **1.5 Dissertation Organization**

The dissertation is organized in six chapters. Chapter 2 reviews of the relevant useful concepts. Chapter 3 proposes a new feature extraction in the classifying of sports type. Chapter 4 presents a new algorithm of blurred background elimination, player area detection, and feature extraction in part of semantic classification. Chapter 5 describes the experimental methods and analysis of experimental results. Chapter 6 summarizes the overall framework and discusses future work.

## CHAPTER II

### RELATED WORK

Image searching is one of the most interests of internet tools because of the benefits for users. Recently, a large digital image including sport images is widely used throughout the internet. Sport image is interpreted based on its contents and textual information. Therefore, developing the searching tools for interpreting the sport images based on their semantic information is a pragmatic and useful work. This chapter some prior works in image classification and video detection in sport area.

#### 2.1 Sport image classification

The general idea is extracting features from the sport image and using pattern recognition techniques to specify the sport class. Some researchers use different kinds of features and classifiers to solve the problems. The following discusses image content classification that reaches the limitations of accuracy because of ambiguities in different sports scenes.

##### 2.1.1 Combination of textual information and image content

Combining technique is useful for webpage image classification by combining two techniques of textual information and image content. Textual information can be gathered from filenames, URLs, page title, and other available webpage information. In 2005, Tian [5] presented a classification system by mixing information from images and textual information surrounding the image in web-based system. This system can handle three sports classes. The features consist of color histograms, color coherence vector, color moments, and autocorrelation as visual features. The visual features are combined with textual information such as Document Object Model (DOM) surrounding the web-image. Researchers also use a context analysis model to combine two different kinds of context information and created a cross-modal correlation analysis and link-based correlation models. The relational model is based on the relational sport vector classifier (RSVC). All these models classify a given image into three different sport classes, namely basketball, football, and baseball.

The shortfalls can occur as the feature extraction and the models description. For example, the anniversary of a company party has a soccer contest so the description is set to be trend anniversary. This cause, erroneous classification of content and textual information technique.

### 2.1.2 Geometrical information

For applications which textual information is not available, the method should rely only on the image context. In 2007, Paluri [1] used only image content in the classification system such as cross-ratio histogram to describe the geometric structure of the playing court of tennis, football, badminton, and basketball. This research could classify sport images even if the camera views change. The researcher used support vector machine (SVM) to be the main classifier but the performance dropped due to a prominent geometrical structures appearing in the image. This shows that there were limitations in using geometrical information for sport classification.

### 2.1.3 Content-based image retrieval

In 2004, Jung [3] used a robust feature set to classify ten sports. The feature combined with color and edge information such as color histogram (CH), color coherence vector (CCV), edge direction histogram (EDH), and edge direction coherence vector (EDCV). The method exploited the inherent characteristics of the image data since different sport images had unique patterns in color and shapes. The sport image was first classified on either indoor or outdoor sport and further classified into ground sports such as water, snow, or other ground sport. The classifier used Bayesian method because it is simple to integrate with multiple features according to class conditions. The result yields relatively low error rate based on a large database.

### 2.1.4 Adaptive of fusion neural network

One approach in solving image classification problem is a modification of a structure of neural network classifier (NNC), called the fusion neural network classification (FNNC) [4]. The fusion function considers the property of equivalent contribution of MPEG-7 descriptors in different dimension features. Figure 2.1 illustrates the structure of FNNC. The architecture is represented by input vector  $x_{jk}$  and output vector  $y_{hi}$ . The sensitivity  $S_{x_{jk}}$  is the sensitivity vector. The weight vector is represented by  $w_{kM}$ . The FNNC is presented having 1 to  $H$  hidden nodes. The outputs from the hidden nodes are represented by  $y_{hi}$ .

Eq. 2.1 is the output function of conventional architecture. Let  $M_1$  and  $M_2$  be the input dimension of  $X_1$  and  $X_2$  such that  $M_1 > M_2$ . The descriptor is defined by  $X$ . The  $X$  descriptors of features  $j$  at dimension 1 to  $M_1$  are defined by  $X_j = [x_{j1}, x_{j2}, \dots, x_{jk}, \dots, x_{jM_1}]$  where  $j$  is 1 or 2. The activation function in hidden layer is given by  $f_h$ . Moreover the corresponding weight is given

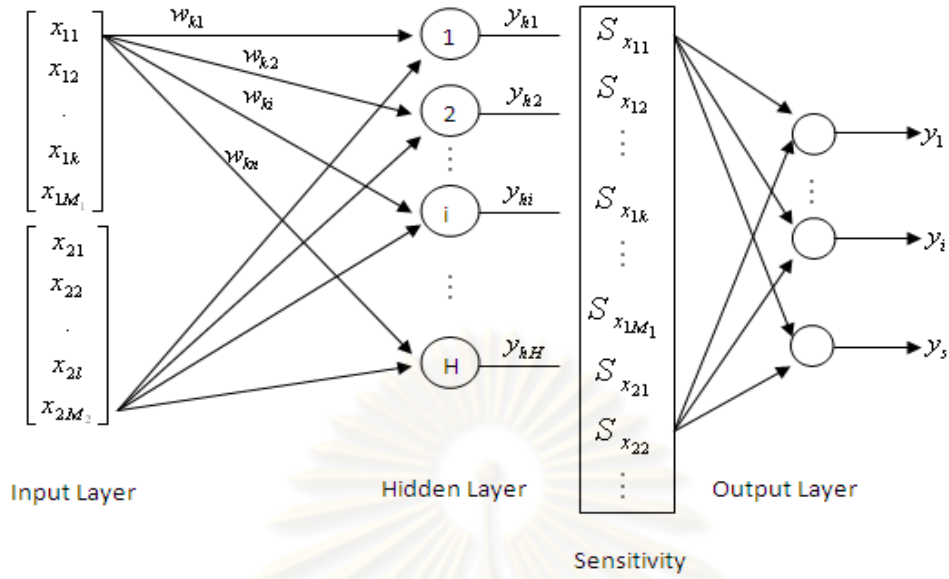


Figure 2.1 The overall architecture of FNNC.

by  $W_{1k}$  and  $W_{2l}$ .

$$y_{h,i} = f_h \left( \sum_{k=1}^{M_1} x_{1k} w_{k,i} + \sum_{l=1}^{M_2} x_{2l} w_{l,i} \right) \quad (2.1)$$

The modified of output function is presented in Eq. 2.2

$$y_{h,i} = f_h \left( \frac{1}{M_1} \sum_{k=1}^{M_1} x_{1k} w_{k,i} + \frac{1}{M_2} \sum_{l=1}^{M_2} x_{2l} w_{l,i} \right) \quad (2.2)$$

The factor,  $1/M_1$  and  $1/M_2$  are average feature values of both descriptors. They are assigned to reduce the contribution of high-different dimensional number. The sensitivity function is represented in Eq. 2.3

$$S_{x_{jk}} = \frac{1}{H} \frac{\partial y_{hi}}{\partial x_{jk}} \quad (2.3)$$

The fusion architecture is more complicated than the formal structure NNC. Although the structure is complicated, the features outperform the performance of sport type classification, i.e., edge histogram (EH) and region-based shape (RS). Consequently, these features are chosen to compare with the proposed features.

## **2.2 Semantic image classification**

A difficult problem is to interpret low-level features to high-level semantics in two dimensional images. Some related works are in the area of still image and video processing.

### **2.2.1 Semantic situation in sport game**

In 2004, Rea [8] detected snooker event in a broadcast video of sports footage. The snooker events are associated with “Break-building”, “Conservative play”, “Escaping a snooker” and “Shot-to-nothing.” Moreover, this work uses Hidden Markov Model (HMM) as a the classifier based on ball tracking method by HSV color model. Tea Mon Bae, et.al [11] presented the detection of situations in golf such as “drive”, “bunker”, “putting” and “walking.” They used a classifier by combining two techniques which were HMM and Support Vector machines (SVM). Although the semantic detection in video can widely detect game situation, the detection methods cannot apply directly to 2-D image.

### **2.2.2 Classifying events by scene and object recognition**

In 2007, Li-Jia Li and Li Fei-Fei [10] proposed a framework to answer the “What, where and who?” in static image under the interpreting of semantic components. They classified the events in the input image with the contribution of a number of semantic object label and scene environments. Eight sport events were used in the experiment and the accuracy was less than 80%. The misclassified object comes from a confusion in the similar foreground objects as bocce and croquet.

In 2008, Kesorn [7] presented a framework to predict sport type of an image which can solve the problem of incomplete text captions. The framework combined ontology of the natural language captions, visual features, color, and edge information. Even though the performance of the results was high but only 240 images in the experiment were not sufficient for testing methodology.

In 2010, Shenghua [9] presented a novel web image based automatic image tagging. The system required users to categorize the input image first and the system would automatically tagged by using spares coding technique and local- global ranking agglomeration. The experimental results outperformed k-NN method. Again, this system used only eight sports in distinguishing of player shape and situation.

To overcome the previous limitations, this dissertation focuses on classifying the similarity of player’s shape under the same kinds of sport.



### 2.2.3 Combined method

Tseng [13] proposed a combined method of data mining and query decomposition techniques, namely Intelligent Concept-Oriented Search (ICOS). This is useful in association with the conceptual object mining and visual ranking of conceptual objects. The conceptual objects associated process consist of object segmentation and annotation process. The public software Normalized Cut and XMLTools were used for the associated process. The idea works on SVM classification at average precision 87.6%. Since this idea has been tested on the images collected from Corel Gallery Professional Image Library and the web search engines on the Internet, totally thirty categories were chosen. The dataset consisted of various categories, so the system did not present the efficiency of sport image classification.

Liu [6] proposed semantic image retrieval based on a novel methodology of decision tree induction, named DT-ST. The system consisted of JSEG image segmentation and machine learning techniques. Each segmented region was extracted by low-level features in a “dominant color” of HSV model and in texture features as the result of Gabor filter. To construct semantic template, a decision tree was used. Although this idea was improved in keyword searching together with region of interested in the image, but the system still classified different objects in the same category. However, the problem of classifying the posture of the same object still not solved.

To reviews about the tools, the details is follows:

## 2.3 Tools reviews

### 2.3.1 YCbCR Color Model

Recently, there are different color models for image representation such as RGB, YCbCr, and HSV color models. Different characteristics are suitable for different color, illumination, and chromatic. In an utilization of image processing, YCbCr and HSV are better than RGB because RGB contains high correlation among R, G, and B components. There are five utilizations of YCbCr color model in recent research, (a) face recognition [14–17], (b) face verification [18], (c) image compression [19], (d) application of YCbCr of identifying the cotton contaminants [20] and (e) color segmentation [21].

- a In face recognition, Lin [14] proposed face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network. [15] proposed a high

performance pose invariant face recognition system on the probability distribution function (PDF). The proposed combine feature vectors obtained from different color channels in HSI and YCbCr color spaces to improve the recognition performance of feature vector fusion (FVF) and majority voting (MV) methods. [17] proposed the utilization of YCbCr transformation to improve color face recognition under various learning scenarios with the chromatic sampling ratios used in MPEG standards. [16] proposed a face detection with complex background based on a mixed skin-color segmentation model in YCbCr and HSI color space.

- b In face verification, [18] proposed face verification on a large scale database with horizontal and vertical 2D principal component analysis (HVDA) by using a color configuration across two color spaces, YIQ and YCbCr. This technique achieves the face verification rate of 78.24% at the false accept rate of 0.1%.
- c In image compression, [19] proposed an effective colorization in YCbCr space on the maximum a posteriori (MAP) estimation of a color image. This technique performs that YCbCr is simpler than RGB space and requires much less computation time.
- d The application of YCbCr is proposed to identify the cotton contaminants. [20] proposed the identification of cotton contaminants using neighborhood gradient based on YCbCr color space. YCbCr color space is useful for distinguish the cotton contaminants with image sharpening algorithm based on neighborhood gradient and threshold segmentation.
- e Especially YCbCr is utilized to color segmentation [21]. [21] proposed color segmentation algorithm used the YCbCr color space with the automatic threshold based on a flame image segmentation and tracking of the process of coal powder burning in the boiler. The proposed supported YCbCr color space is better to discriminating the luminance from the chrominance which is robust to illumination changes than RGB color space.

From these researches, it can be concluded that YCbCr is suitable for color segmentation in discriminating the luminance.

### 2.3.2 HSV Color Model

HSV (Hue, Saturation, and Value/brightness) is the color model which decomposed hue and saturation which correspond to human perception and valuably for image enhancement. The useful of HSV is enhancement of skin color detection [22–24] application of head detection algorithm [25], and application of identifying computer graphics [26].

In details of face detection, [22] proposed enhancement of skin color detection with the analyzing the distribution of skin color in HSV color space. [23] proposed an automatically method for the frontal human face region location under complex background in color image. The method utilizes color histogram of YCbCr and HSV color model to detect the skin area. [24] proposed a novel face detection and skin segmentation using Takagi-Sugeno (T-S) fuzzy model and HSV color model. The detection model uses fuzzy classifier in conjunction with HSV color model to quickly locate face area in the image. In other words, HSV outperform to faces and skin detections. [25] proposed a new head detection algorithm capable to handling significantly variable conditions in terms of viewpoint. The proposed uses Gaussian mixtures to detect the appearance of color distribution models of hair and skin based on XYZ and HSV color spaces.

The details of application of HSV color mode, [26] proposed a novel approach to distinguish computer graphics from photographic image with statistical moments and wavelet subbands. The proposed utilization of HSV color space with decoupled brightness from chromatic components demonstrated better performance than RGB color model.

HSV is very useful for face and skin detection. Moreover the application is usability for decoupling brightness from chromatic components which are significant for player area in sport image. The foreground object such as the player area presented more brightness than its background. Consequently, these studies choose HSV as the considered plane in blurred background elimination algorithm.

### **2.3.3 Discrete Cosine Transform (DCT) technique**

DCT is the widely used tools to transform the spatial domain to frequency domain. DCT is an important tool in numerous applications of image processing in terms of lossy compression, object recognition [27, 28], and application of face recognition [29].

There are many previous studies of utilization of DCT to improve the performance of object recognition. First, [27] proposed the utilized of DCT coefficient to improve the performance of object recognition with luminance, rotation, and location invariance. The performance of luminance and rotation invariance is illustrated by reducing error rates in face recognition using a two-dimensional DCT proposed on image synthesis procedure. [28] proposed a new algorithm for detecting and tracking multiple moving object in both outdoor and indoor environments with the measuring method of the change of feature vector. The method utilizes DCT coefficient to extract the texture of object and combine color-texture in each block to be the feature vector. The method represented the multiple cues including color, texture, and spatial position based on object tracking with inexact graph

matching. [29] proposed the useful of low frequency components of DCT and Principal Component Analysis (PCA) to solved the illumination invariant in face recognition. The illuminated image is normalized by low frequency components of DCT.

Consequently, DCT technique is suitable for object tracking and recognition.

#### **2.3.4 Discrete Wavelet Transform (DWT) technique**

2D Discrete Wavelet Transformation (DWT) is widely used in multiresolution of image enhancement, i.e., image compressing [30], effective watermarking [31], and image enhancement [32].

[30] presents the standard tool of DWT in JPEG2000 compression and more in efficient of decomposition and reconstruction for image coding. The proposed method presents the useful of wavelet decomposition in high- and low-decomposition.

[31] proposed an effective way to improve digital watermarking based on Singular Value Decomposition (SVD) in DWT domain. The method is transformed the original image to DWT and SVD by Arnorld transform.

[32] proposed a new satellite image by enhancing the contrast with the DWT and SVD. The method used the decomposition of DWT in four frequency subbands and applied SVD to estimate low subband image for reconstruction of DWT's inverse.

From the previous studies, DWT is effective in image decomposition and reconstruction. In addition DWT is efficient to encode the image contents. It can enhance the contrast of the input image based on subbands reconstruction. In the other words, DWT is suitable for decomposing high- and low-contrast, which is important content in blurred area removal.

#### **2.3.5 2D Discrete Fourier Transform (2D-DFT)**

The fundamental of frequency transformation is Fourier transform. For digital image is applied by 2D DFT. The DFT converts the spatial image into components of different frequencies. Fast Fourier Transform (FFT), a version of DFT, attempts to minimize computational time of transformation. There are many applications of FFT, i.e., object recognition [33–35], and [36]

[33] proposed a framework for scaling, translating and rotating object using the short time Fourier transform. The short time section is used to be the features which is useful for discriminating variants of similar objects. The application of Fourier transformation is efficient for computation and robustness in the presence of noise. [34] proposed object recognition based on Fourier descriptors under the different circumstances such as noise and similarly object. [35] proposed the effective tools using Fourier descriptors and Genetic algorithm for object recognition. The method used Fourier

descriptor for feature representation and used genetic algorithm technique to map the object features. Finally, [36] proposed multi-feature to improve the performance of object class recognition. This input comes from Fourier and Scale-Invariant Feature Transform (SIFT) Descriptors. However, only SIFT descriptors is not enough. The combined features outperform better than using only SIFT descriptors up to 15% with limitation of views of images.

Therefore Fourier transform is a useful tool for object recognition with invariance of scaling, translation, and rotation.

### 2.3.6 Singular Value Decomposition (SVD)

SVD is a linear algebra tool of factorization for a real or complex matrix. It widely uses in signal and image processing. SVD is useful in image compression technique [37–39] and watermarking technique [40] based on the effective matrix compaction.

[37] proposed the utilization of SVD to compress and reduce the storage space on images. The method uses SVD properties of energy compaction adapting to the local statistical variations of image matrix.

[38] proposed the hybrid compression between DWT and SVD to exploit the best energy compaction for monochrome images. The method is processed by the decomposing of sub-blocks of an original image applied with 2D wavelet transform and SVD decomposition, obtained maximum energy of the image in terms of a minimum number of coefficients.

[39] investigated performance of DCT in image quantization and SVD decomposition based on the ranking of coefficient in image compression. The numerical analysis is measured in compression ratio.

[40] proposed watermarking technique with frequency domain and SVD. The method used DCT based watermarking techniques for compression and DWT based compression for scalability. Consequently, combination tools of DWT, DCT and SVD are proposed for non-blind transform domain watermarking based on the embed DWT coefficients of watermarking information by DCT coefficients.

For the conclusion, SVD is suitable for decomposing the information hiding in the image based on frequency domain.

All relevant researches did not aim to recognize posture but classify only scene environment. Six sports, namely, baseball, basketball, field and track, soccer, skiing, and swimming are considered in this dissertation. The semantic query will involve the posture of player, number of players, color of clothes, sport equipment, and ground of sport. These six sports consist of three background types



that are water sports, field sports, and ice sports. Details will be discussed in Chapter 4.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## CHAPTER III

### SPORTS IMAGE CLASSIFICATION

Sports image classification is the first process in semantic classification process. The accuracy is not high but the complexity of algorithms is high. To overcome constrains of the complexity, this dissertation introduces simplified features with background-avoided extraction process. These new features are more descriptive to the player's posture. The technique can select similar postures and background texture better than other techniques [4] relying on low-level image features. There are four features investigated that are 64-bins color histogram, DCT coefficient, and Cb and Cr color models descriptors. To improve the posture recognition, Cb and Cr color spaces are employed for player region segmentation. The other two features are used for representing the color and texture of ground sports. In addition, singular value decomposition (SVD) is applied to decompose large feature matrix into input vectors of acceptable length.

#### 3.1 Majority Color Extraction

Some sport images are rather easy to recognize by majority colors because of the dominant color of ground. For example, blue or light blue color is the majority of swimming images while white color is the majority of skiing images. Hence, majority colors can describe type of sport. Given each sport image is in RGB and grayscale formats defined by  $I(m, n)$  and  $I_g(m, n)$ , respectively. At any pixel, the intensity of each color is encoded by eight bits. Let  $r_i$ ,  $g_i$ , and  $b_i$  be the intensity of red, green, and blue at the pixel, respectively. The value of each color intensity is represented as follows:

$$r_i = (r_{i,1}, r_{i,2}, \dots, r_{i,8}) \quad (3.1)$$

$$g_i = (g_{i,1}, g_{i,2}, \dots, g_{i,8}) \quad (3.2)$$

$$b_i = (b_{i,1}, b_{i,2}, \dots, b_{i,8}) \quad (3.3)$$

Here,  $r_{i,1}$ ,  $g_{i,1}$ , and  $b_{i,1}$  are most significant bits. To capture the informative color at any pixel  $i$ , only the first two significant bits of each color are selected and concatenated to form a 6-bit majority color value, denoted by  $c_i$ . The value is significant enough to be used as a majority

Table 3.1 An example of majority colors calculation for field, track, ice, and water backgrounds.

Background Types	R,G, and B values in decimal and binary digits		6-bits majority color	
	Decimal	Binary	Binary	Decimal
Field(Green)	85:106:50	01010101:01101010:00110010	10100	20
Track(Orange)	170:83:73	10101010:01010011:01001001	100101	37
Ice(White)	202:201:207	11001010:11001001:11001111	111111	63
Water(Blue)	54:162:211	00110110:10100010:11010011	1011	11

color feature of the image. Therefore, at pixel  $i$  the majority color value  $c_i$  is represented as  $c_i = (r_{i,1}, r_{i,2}, g_{i,1}, g_{i,2}, b_{i,1}, b_{i,2})$ . There are 64 values which can be viewed as 64 bins.

Table 3.1 illustrates the calculating process of the majority color values of some pixels for track and field, ice, and water backgrounds. The color histogram is created from the corresponding color value of 64-bin in decimal value for all pixels.

### 3.2 Descriptor for Complex Background

A complex background is different from a normal background in terms of number of colors, intensity, and distribution of color pixels. To capture background characteristics, the distribution of colors with different intensities must be transformed to frequency domain. High frequencies are presented on the different value of intensity between two adjacent pixels. On the other hand, low frequency is the low value of the difference between two adjacent pixels. Edge information in the input image will be decomposed by discrete cosine transform (DCT) [41, 42]. Therefore, the appropriate descriptor to capture any complex background can be efficiently represented by using the coefficients derived from DCT.

The process transforms an RGB color image into gray-leveled format. Then, two-dimensional discrete cosine transform is applied to compute all coefficients used as the descriptor of the image. 2D-DCT basis function is the decorrelation method transformed into sub-sequence coefficients [43]. 2D-DCT at position  $x$  and  $y$ ,  $D(x, y)$ , can be calculated from the equation 3.4. Denote  $x$  and  $y$  to be the coordinate indices of coefficient of 2D-DCT. The calculation is as follows.

$$D(x, y) = \sigma(x) \sigma(y) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} g(m, n) \cos \left[ \frac{\pi(2m+1)x}{2M} \right] \cos \left[ \frac{\pi(2n+1)y}{2N} \right] \quad (3.4)$$

$$\sigma(x) = \begin{cases} \sqrt{\frac{1}{M}} & \text{for } x = 0 \\ \sqrt{\frac{2}{M}} & \text{for } x \neq 0 \end{cases}$$

where  $x = 0, 1, 2, \dots, M - 1$  and  $y = 0, 1, 2, \dots, N - 1$

The formulation is calculated on the intensity of gray-level image. Let function  $g()$  be the matrices function of grayscale intensity of input image and the grayscale intensity at indices of  $m$  and  $n$  be represented by  $g(m, n)$ .

### 3.3 Posture Descriptor

Postures of a player can be directly recognized from the outlined shape of the player. To increase the difference between player foreground and complex background, the RGB format is transformed to YCbCr format by the following equations [44].

$$Y = 0.0299R + 0.587G + 0.114B \quad (3.5)$$

$$Cb = 0.564(B - Y) + 128 \quad (3.6)$$

$$Cr = 0.713(R - Y) + 128 \quad (3.7)$$

The advantage of YCbCr is presented on face detection in complicate background and different illumination [14] and JPEG-compressed color images is applied [45].

The high performance of YCbCr transformation in sport image is supported by the colorful of player's suit. The performance decreased in direct variation to the decreasing of illumination value. The objective of this subject is highlighting the player area by coloring technique. Consequently, if the player's suit has similar blue or red color, the result of color highlighting technique will not work. For example, Figure 3.1 shows the result of blue player's cloths. The Cb transformation (in the top row of the Figure 3.1) clearly highlights the player area because of the illumination of foreground object. However, other insensitive cases such as black color which lacks lighting information. YCbCr transformation of black player's suit is presented in the bottom row of Figure 3.1.

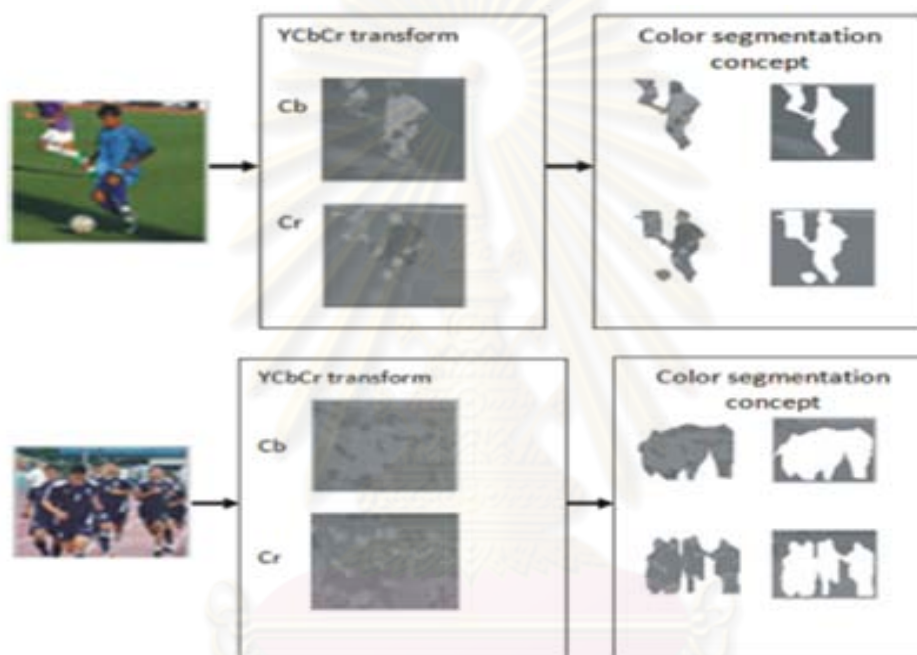


Figure 3.1: Examples of transforming RGB color format to YCrCb color format and the comparison of outstanding shapes from Cb and Cr color space. Top row: an example of blue color image denoted as the result of Cb or blue-difference value. Bottom row: an example of Cb and Cr color spaces for black clothes.



### 3.4 SVD on image representation concept

After extracting the image feature, the process of data dimension reduction is considered. A proper tool of reducing a high dimensional to a lower dimensional spaces is singular value decomposition tools (SVD) [46]. The SVD tool is well known in signal processing and statistics applications. Consequently, this dissertation uses SVD to factorize a complex matrix of image contents such as player's shape in Cb and Cr matrices. The result of factorization is a principal component matrix which is ordered from the most to the least variation, in diagonal matrix.

The complex background and posture descriptors are represented in three matrices, i.e., DCT, Cb, and Cr. The size of each matrix is equal to the size of the input image. The high dimensional of input features complicate the training process of neural network architecture. Reducing the high dimension of feature matrices to low dimensional input vector is required. Consequently, SVD is used to decompose each feature matrix of DCT, Cb, and Cr into three matrices. The resulting matrices are diagonal matrices  $S$ , shown by equation 3.8. The input vector is performed from the diagonal value of matrix  $S$ .

Let  $U_{[M \times N]}$  be a matrix of size  $M \times N$ . Matrix  $U_{[M \times N]}$  can be decomposed by the following equation.

$$U_{[M \times N]} = X_{[M \times M]} S_{[M \times N]} Y_{[N \times N]}^T \quad (3.8)$$

$X_{[M \times M]}$  and  $Y_{[N \times N]}$  are orthogonal matrices such that  $X_{[M \times M]}^T X_{[M \times M]} = I$  and  $Y_{[N \times N]}^T Y_{[N \times N]} = I$ . The columns of  $X_{[M \times M]}$  are orthonormal eigenvectors of  $U_{[M \times N]} U_{[M \times N]}^T$  and the columns of  $Y_{[N \times N]}$  are orthonormal eigenvectors of  $X_{[M \times M]}^T U_{[M \times N]}$ .  $S_{[M \times N]}$  is a diagonal matrix containing the square roots of eigenvalues from  $X_{[M \times M]}$  and  $Y_{[N \times N]}$  in descending order. After applying SVD to DCT, Cb, and Cr matrices, all diagonal elements in each corresponding matrix  $S_{[M \times N]}$  is used as features of each DCT, Cb, and Cr matrix. The illustration of features structure in the system is shown in Figure 3.2 and features array structure is presented in Figure 3.3.

The input image of RGB color image is of size 130. The feature vector consists of SVD matrices having DCT, Cb, and Cr which are  $130 \times 3$  or 390 dimensions. Another additional feature is 64-bins color histogram which increases the feature dimensions to  $390 + 64 = 454$ . This feature vector is presented by a matrix of size  $454 \times 1$  in Figure 3.2. All input vectors are fed to an Artificial Neural Network Classifier (ANN).

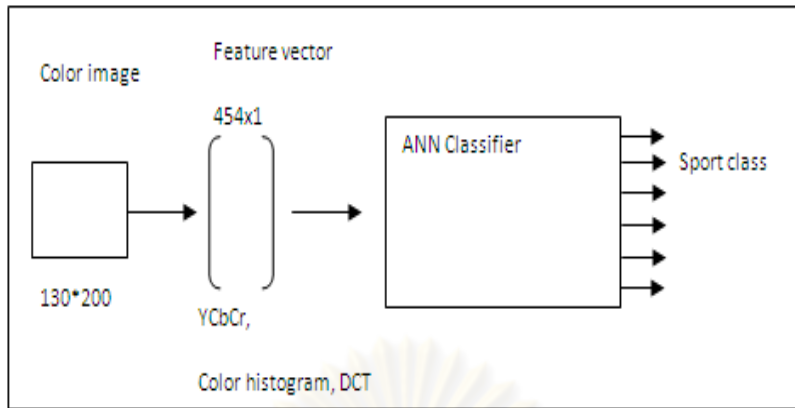


Figure 3.2 Overall architecture illustrated on the feature vector.

Feature structure of SVD decomposition and its diagonal result are illustrated in Figure 3.3. The diagonal elements of feature matrix  $i$  at position  $m$  and  $n$  is defined by  $\sigma_{m,n}^i$ .

### 3.5 Algorithm of sport image classification

For overall architecture in Figure 3.4, the classification system uses the feature vector as the input of ANN. The images are resized to the average size of  $130 \times 200$  pixels because of suitable dimensions of input feature for neural network classifier. The feature vector is extracted and arranged in a column vector of 454 dimensions as shown in Figure 3.2. There are two layers of ANN architectures, hidden layer, and output layer. The output node consists of six bits corresponding to six kinds of sports.

Define intensity components of RGB image by  $I(m, n)$  at coordinate  $(m, n)$  of size  $[M \times N]$ . Let intensity of grayscale image be  $I_g(m, n)$ . To define fundamental algorithm for decompose the image contents.

**Algorithm: Singular Value Decomposition** Define the considered matrix of size  $[M \times N]$  as  $U_{[M \times N]}$  and decomposed to three matrices  $X_{[M \times M]}$ ,  $S_{[M \times N]}$ , and  $Y_{[N \times N]}^T$ . The diagonal singular values are represented by  $S_{[M \times N]}$ . The decomposition is calculated by the following equation:

$$U_{[M \times N]} = X_{[M \times M]} S_{[M \times N]} Y_{[N \times N]}^T \quad (3.9)$$

**Algorithm: YCbCr transformation** The input image is the RGB image of size  $[M \times N]$ . Y, Cb, and Cr are calculated from the following equations:

### S-Cb and Cr

$$\begin{pmatrix} 130 \times 200 \end{pmatrix} \xrightarrow{\text{SVD}} \begin{bmatrix} \sigma_{1,1}^1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \sigma_{130,130}^1 & 0 \end{bmatrix}_{130 \times 200} \Rightarrow \begin{bmatrix} \sigma_{1,1}^1 \\ \sigma_{1,2}^1 \\ \vdots \\ \sigma_{1,130}^1 \end{bmatrix}$$

### S-DCT

$$\begin{pmatrix} 130 \times 200 \end{pmatrix} \xrightarrow{\text{SVD}} \begin{bmatrix} \sigma_{1,1}^3 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \sigma_{130,130}^3 & 0 \end{bmatrix}_{130 \times 200} \Rightarrow \begin{bmatrix} \sigma_{1,1}^3 \\ \sigma_{1,2}^3 \\ \vdots \\ \sigma_{1,130}^3 \end{bmatrix}$$

Figure 3.3 Performed input vectors from SVD matrices on S-Cb, S-Cr and S-DCT.

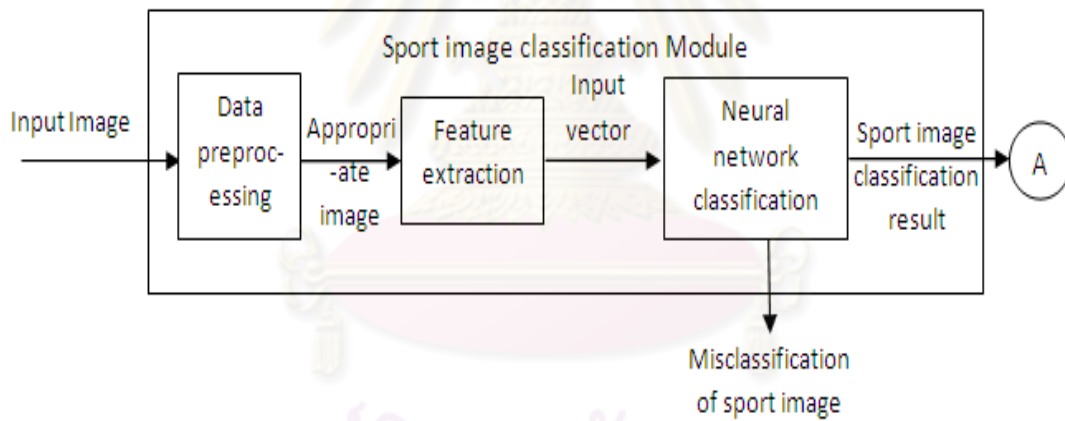


Figure 3.4 The overall system of sport image classification module.

$$Y = 0.0299R + 0.587G + 0.114B \quad (3.10)$$

$$Cb = 0.564(B - Y) + 128 \quad (3.11)$$

$$Cr = 0.713(R - Y) + 128 \quad (3.12)$$

The preprocessing image is resize all images to  $[130 \times 200]$  for the suitable dimensions of input features for neural network classifier. The overall process of the sport image classification module is separated into eight processes given in the Algorithm Sport image classification.

**Algorithm: Sport image classification**

- 1: Calculate 64-bins color for each pixel  $i$  from red ( $r_i$ ), green ( $g_i$ ), and blue ( $b_i$ ) which the intensity colors encoded by eight bits.

$$r_i = (r_{i,1}, r_{i,2}, \dots, r_{i,8}) \quad (3.13)$$

$$g_i = (g_{i,1}, g_{i,2}, \dots, g_{i,8}) \quad (3.14)$$

$$b_i = (b_{i,1}, b_{i,2}, \dots, b_{i,8}) \quad (3.15)$$

$$c_i = (r_{i,1}, r_{i,2}, g_{i,1}, g_{i,2}, b_{i,1}, b_{i,2}) \quad (3.16)$$

- 2: Calculate histogram of 64-bins color value.

$$64\text{bins\_hist}(j) = \sum_{i=1, c_i=j}^{M \times N} 1 \quad (3.17)$$

- 3: Transform RGB image to grayscale image represented by  $I_g(m, n)$ .
- 4: Apply 2D-DCT to the transformed image to achieve corresponding 2D-DCT coefficients by the following equation.

$$D(x, y) = \sigma(x)\sigma(y) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_g(m, n) \cos \left[ \frac{\pi(2m+1)x}{2M} \right] \cos \left[ \frac{\pi(2n+1)y}{2N} \right] \quad (3.18)$$

$$\sigma(x) = \begin{cases} \sqrt{\frac{1}{M}} & \text{for } x = 0 \\ \sqrt{\frac{2}{M}} & \text{for } x \neq 0 \end{cases} \quad (3.19)$$

where  $x = 0, 1, 2, \dots, M - 1$  and  $y = 0, 1, 2, \dots, N - 1$

- 5: Decompose 2D-DCT coefficient matrix from previous step by algorithm of decompose SVD coefficient.
- 6: Transform RGB image to YCbCr domain by algorithm of YCbCr transformation.
- 7: Decompose Cb and Cr matrices by algorithm of SVD decomposition. The diagonal matrix defined by  $S_{[M \times N]}$
- 8: Concatenate the diagonal elements of matrices  $S_{[M \times N]}^{DCT}$ ,  $S_{[M \times N]}^{Cb}$ , and  $S_{[M \times N]}^{Cr}$ , and 64-bins color histogram to form a feature vector.

### 3.5.1 Examples of image features

Examples of features extraction are illustrated in the following figures. 64-bins color histogram features of six images from six sports are shown in Figure 3.5. One sub-figure consists of an original image on the left and 64-bins color histogram on the right. The six sports from left to right and top to bottom are baseball, basketball, field & track, skiing, and swimming.

The second feature is YCbCr color model as illustrated in Figure 3.6. One sub-figure presented four sub-images of the original image (on the top left), Y color mode (on the top right), Cb color mode (on the bottom left), and Cr color mode (on the bottom right). Six images from six sports are presented in YCbCr color transforms as shown in the Figure 3.6.

The last feature S-DCT is shown in Figure 3.7. One sub-figure is shown in the original image and its S-DCT graph. Example images of six sports are presented in this Figure.



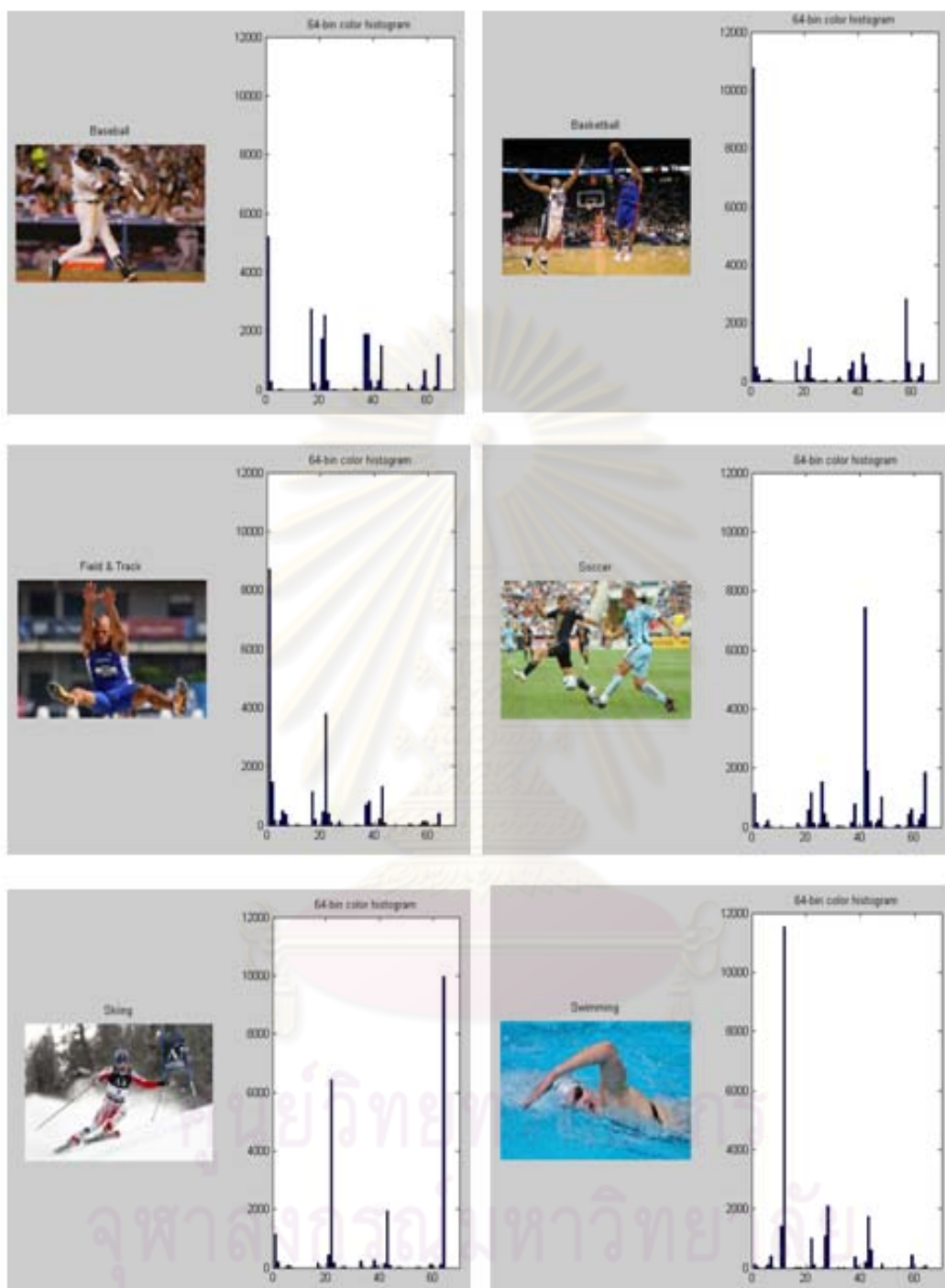


Figure 3.5: Example of 64-bins color histogram feature in six sports. The left column is input image and right column is 64-bins color histogram. Top row: baseball and basketball. Middle row: track & field and soccer. Bottom row: skiing and swimming.



Figure 3.6: Example of YCbCr feature in six sports. The left-top sub-image is the input image and the right-top sub-image is Y color mode. The bottom-left is Cb color mode and bottom-right is Cr color mode. In top row illustrate images from baseball and basketball, in middle row presents images from field & track and soccer, and in bottom row shows images from skiing and swimming.

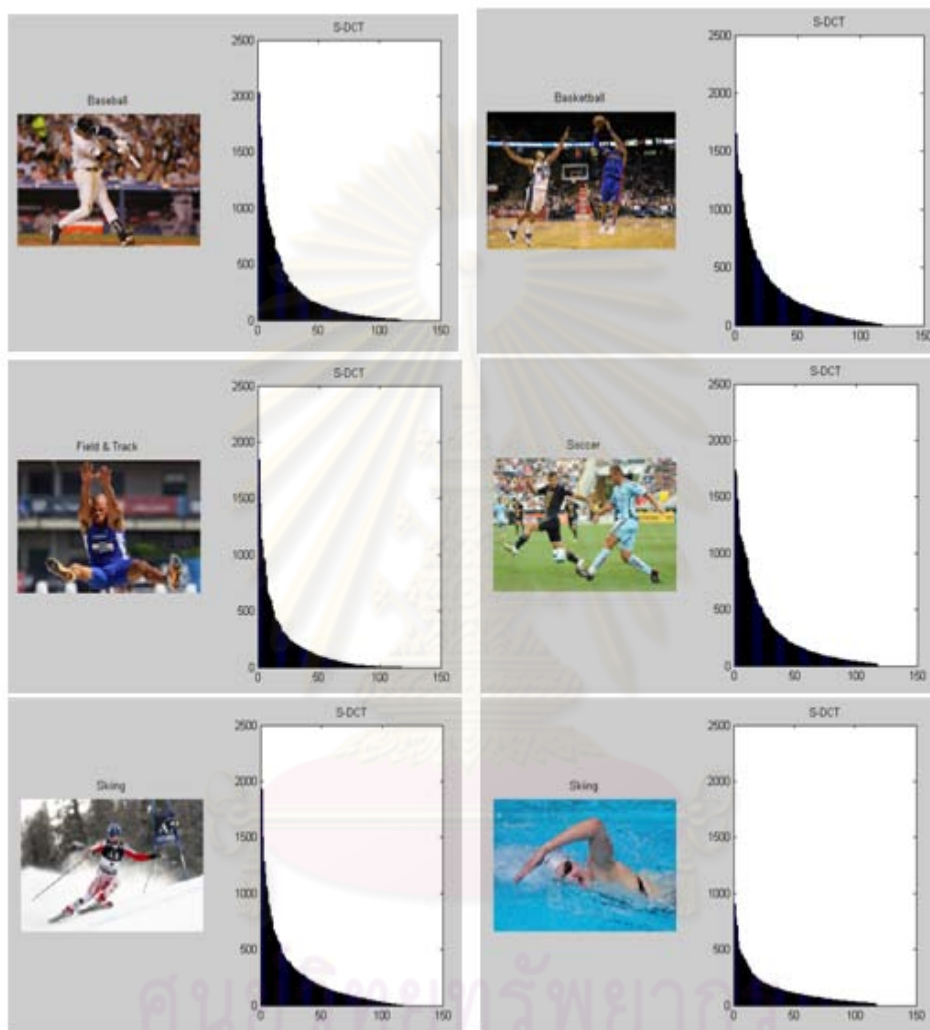


Figure 3.7: Example of S-DCT feature in six sports. The left column of sub-image is the input image and the right column of sub-image is S-DCT. In top row illustrate images from baseball and basketball, in middle row presents images from field & track and soccer, and in bottom row shows images from skiing and swimming.

## CHAPTER IV

### POSTURE CLASSIFICATION

There are two regions of sport image that must be considered: focus and out-of-focus regions. Normally, the player area is the focused area and the background is the out-of-focus area. The out-of-focus area has the properties of low intensity in color and more brightness than the focused area. Low contrast value and high contrast value can be separated by the different value between two adjacent pixels. One tool approach to detect the different value between two adjacent pixels is in the frequency transformation. The 2D Discrete Wavelet Transform (2D-DWT) decomposes the different color intensity along three directions; horizontal, vertical and diagonal directions. In this study, only diagonal direction is used to detect the abnormal curve of player shape, which is better than detecting in horizontal or vertical directions. Blurred background elimination is processed on the intensity value in V channel of HSV color model. Then, blurred background is removed by a high pass filtering based on 2D-DWT and its inverse with basis 'Haar' wavelet function.

The framework is separated to four parts. Four proposed steps compose of blurred background elimination, irrelevant area elimination, player area elimination, and semantic feature extraction.

#### 4.1 Blurred background elimination

In sport image, the obvious blurred background is important to support posture recognition. For the biggest field such as baseball or soccer field, the blurred background area is usually presented in the sport image. Blurred background area should be removed at the first of all process. The sport image is separated into two following major regions.

- Focused region is a foreground object or a player area which is the region of interest. Define this area as the *high-contrast* region which has the properties of high intensity in color and brightness.
- Out-of-focus region is the background area. Define this area as low-contrast region which has *low intensity* in color and brightness.

The contrast property is the different value between two adjacent pixels. The best tool to detect the differential of contrast property is wavelet transformation. In general, the decomposition of 2D-

DWT is done in three directions, namely horizontal, vertical and diagonal directions. In this study, only diagonal direction is used to detect the abnormal curve of player shape, which is better than detecting in horizontal or vertical directions. Blurred background elimination is processed on the intensity value in V channel of HSV color model. Then, blurred background is removed by a high pass filter using 2D-DWT and its inverse based on 'Haar' wavelet.

The result from the process of high pass filter contains some noisy data needed to remove by applying a threshold value. The threshold value is determined by the gradient magnitude of the result image. The gradient magnitude is the different value of intensity between two adjacent pixels. In this dissertation, 5% of the gradient value is used as a threshold value for field & track and swimming, and 10% for the others sports because of their characteristics the of image view.

From the previous process, the player boundary pixels are extracted. The post-process after that is applying dilation technique for connecting the boundary and fulfilling hold area in the player area. Two kernel functions of the dilation are ball shape and line shape. Since the detection of wavelet transformation corresponds to diagonal direction, the dilation is applied to 45 degrees and 90 degrees. In addition, the length of line kernel is normally defined by 5 pixels but only field & track is 7 pixels used because the results of background elimination in track floor background are few. The details of the Algorithm is illustrated as follows.

**Algorithm: Blurred background elimination**

- 1: Create a new image  $W$  such that each pixel value at position is  $(m, n)$  defined as follows:

$$W(m, n) = \frac{1}{3}(r(m, n) + g(m, n) + b(m, n)); \quad 1 \leq m \leq M; \quad 1 \leq n \leq N \quad (4.1)$$

where  $r(m, n)$ ,  $g(m, n)$ , and  $b(m, n)$  are intensity of red, green, and blue components.

- 2: Apply Haar wavelet transform to image  $W$  to obtain the diagonal component in a form of matrix  $D$ .
- 3: Apply inverse Haar wavelet transform to matrix  $D$ .

$$W'(m, n) = IDWT(D) \quad (4.2)$$

where  $IDWT(\cdot)$  is the function of inverse discrete wavelet transform.

- 4: Reset each pixel value  $W'(m, n)$  at position  $(m, n)$  of matrix  $W'$  as follows:

$$W'(m, n) = \begin{cases} W'(m, n) & \text{if } W'(m, n) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.3)$$

- 5: Repeat steps 2 - 4 three more times to denoise image.



6: **for** each pixel coordinate  $(m, n)$  **do**

7: Calculate gradients in x-y directions denoted as  $G_x(m, n)$  and  $G_y(m, n)$  by the following equations.

$$G_x(m, n) = W(m, n + 1) - W(m, n - 1) \quad (4.4)$$

$$G_y(m, n) = W(m + 1, n) - W(m - 1, n) \quad (4.5)$$

8: Calculate the actual gradient  $G(m, n)$  at position  $(m, n)$  as follows:

$$G(m, n) = (G_x^2(m, n) + G_y^2(m, n))^{1/2} \quad (4.6)$$

9: **end for**

10: Calculate threshold for each sport by:

$$T = r \cdot [\max_{\forall m, n}(G(m, n)) - \min_{\forall m, n}(G(m, n))] + \max_{\forall m, n}(G(m, n)) \quad (4.7)$$

where  $r$  is 0.1 for all sports but 0.05 for field & track and swimming.

11: Create an image for  $W''$  for dilation such that each pixel value  $W''(M, N)$  at position  $(M, N)$  is defined as follows:

$$W''(m, n) = \begin{cases} I_g(m, n) & \text{if } W'(m, n) > T \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

12: Apply dilation to the image  $W''$  with ellipsoid shape kernels whose radius and height are 2 and 4 pixels, respectively.

13: Fill holes of image  $W''$ .

14: Apply dilation to image  $W''$  (for each sport except field and track) by using line shape kernel of 5 pixels in 45 degree with respect to horizontal axis. In case of field and track, do the same step but the initial length of kernel is set to be 7 pixels.

15: Apply dilation to image  $W''$  (for each sport except field and track) by using line shape kernel of 5 pixels in 90 degree with respect to horizontal axis. In case of field and track, do the same step but the initial length of kernel is set to be 7 pixels.

16: Repeat steps 14 and 15 by decreasing the length of kernel by 1 until the length of last kernel is 1 pixel.

17: Apply dilation to image  $W''$  with ellipse shape kernels whose horizontal and vertical radii are 3 and 2 pixels, respectively.

18: Fill hole of image  $W''$ .



19: Adjust the intensity of image  $W''$  to black and white image at entry  $(m, n)$  calculated by the set of:

$$W''(m, n) = \begin{cases} 1 & \text{if } W''(m, n) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

20: Select the largest connected region of image  $W''$ .

21: Restore all pixels in the largest connected region to gray-level intensity from original gray image  $I_g$  by the following equation and name new image as image  $R$ .

$$R(m, n) = \begin{cases} I_g(m, n) & \text{if } W''(m, n) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

An example of blurred background elimination is presented in Figure 4.1. The sub-image (a) is the original image and sub-image (b) is the value mode of HSV color transformation. The sub-image (c) illustrates the results of 2D-DWT elimination process. The sub-image (d) shows the applied thresholding. In sub-image (e) is the grayscale image from the increasing of intensity. Sub-image (f) is the result from ball shape dilation whose radius and height are 2 and 4 pixels. Sub-image (g) is the line shape dilation. Sub-image (h) presents the result from ball shape dilation of whose radius and height are 3 and 2 pixels. Sub-image (i) is the biggest area of pixels connected. Finally, the result image of the algorithm shows in grayscale image in sub-image(g).

## 4.2 Proposed irrelevant area elimination

The result from the blurred background elimination is not clear in player area. Irrelevant pixels or ground pixels under the player area are irrelevant area. The definition of each area in sport image presented in Figure 4.3. The Figure 4.3 consist of five sub-image ; (a) original image, (b) blurred area, (c) player area, (d) high-contrast area, and (e) low-contrast area. The irrelevant area according to high and low contrasts will be removed by the following process.

### Algorithm : Irrelevant area elimination

#### (1) High-contrast area elimination

1: Create a new image  $\bar{I}$  from image  $R$  in algorithm Blurred background elimination with mean value as follows:

$$\bar{I}(m, n) = R(m, n) - \frac{1}{N} \sum_{n=1}^N R(m, n) \quad 1 \leq n \leq N \quad (4.11)$$

2: **for** each pixel coordinate  $(m, n)$  **do**

3: Calculate gradients in x-y directions denoted as  $G_x(m, n)$  and  $G_y(m, n)$  by the following equations:

$$G_x(m, n) = W(m, n + 1) - W(m, n - 1) \quad (4.12)$$

$$G_y(m, n) = W(m + 1, n) - W(m - 1, n) \quad (4.13)$$

4: Calculate the actual gradient  $G(m, n)$  at position  $(m, n)$  as follows:

$$G(m, n) = (G_x^2(m, n) + G_y^2(m, n))^{1/2} \quad (4.14)$$

5: **end for**

6: Calculate threshold for each sport by:

$$T = r \cdot [\max_{\forall m, n}(G(m, n)) - \min_{\forall m, n}(G(m, n))] + \max_{\forall m, n}(G(m, n)) \quad (4.15)$$

where  $r$  is 0.35 for all sports but 0.08 for swimming.

7: Create an image  $H$  for such that each pixel value  $H(m, n)$  at position  $(m, n)$  is defined as following:

$$H(m, n) = \begin{cases} |\bar{I}(m, n)| & \text{if } |\bar{I}(m, n)| > T \\ 0 & \text{otherwise} \end{cases} \quad (4.16)$$

8: Dilate all pixels of image  $H$  by a median filtering with 3-by-3 neighborhood.

## (2) Low-contrast area elimination

1: Create the new image  $L$  of size  $(m, n)$  form the result image from high-contrast area elimination calculated to RGB image as following:

$$L(m, n) = \begin{cases} I(m, n) & \text{if } H(m, n) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.17)$$

2: Set new value of each  $L(m, n)$  equal to  $\left| L(m, n) - \frac{1}{N} \sum_{n=1}^N L(m, n) - \frac{1}{M/2} \sum_{m=M/2}^M L(m, n) \right|$ .

An example of irrelevant area elimination process is illustrated step by step in Figure 4.2. High contrast elimination process is in the top row of the Figure. Sub-image (a) is the gray-level image which is the result image from blurred background elimination process. Sub-image (b) is the result from Algorithm *Low-contrast area elimination* in step 1. Sub-image (c) is the result of applied the

threshold value from the result image from the above step. For this process, the irrelevant area is removed. Sub-image (d) is RGB intensity based on the remaining pixels having value more than zero. Low-contrast area elimination process is the process to remove the blurred area which presents between player legs. Sub-image (e) shows the result of eliminating means value in each row. Sub-image (f) presents the result of the dilation from Algorithm *low-contrast area elimination* in step 4. sub-image (g)The result image which present the region of interest in the final step.



### 4.2.1 Proposed player area selection

Player area is selected from the center of player area which have the highest of pixels histogram along with row index or column index. The player area is selected by normal distribution which corresponds to the pixel histogram in horizontal and vertical directions. The result is illustrated in Figure 4.4.

#### Algorithm: Player area selection

- 1: Create the new image  $I_{bw}$  for such that each pixel value  $L(m, n)$  at position  $(m, n)$  is defined as following:

$$I_{bw}(m, n) = \begin{cases} L(m, n) & \text{if } L(m, n) > T \\ 0 & \text{otherwise} \end{cases} \quad (4.18)$$

- 2: Calculate pixel histogram in x- and y-direction as follows:

$$h_x(n) = \sum_{m=1}^M I_{bw}(m, n); \quad 1 \leq m \leq M; \quad 1 \leq n \leq N \quad (4.19)$$

$$h_y(m) = \sum_{n=1}^N I_{bw}(m, n); \quad 1 \leq m \leq M; \quad 1 \leq n \leq N \quad (4.20)$$

- 3: Calculate index of maximum histogram pixel histogram along with x-direction  $h_x$  to be the middle index of normalize distribution technique.

$$mid_x = \arg \max_n (h_x(n)) \quad (4.21)$$

- 4: Calculate summation pixels in the histogram:

$$sum = \sum_{n=1}^N h_x(n) \quad (4.22)$$

- 5: Calculate range of extension index

$$i_{ext} = \begin{cases} M - mid_x & \text{if } mid_x \leq M/2 \\ mid_x & \text{if } mid_x > M/2 \end{cases} \quad (4.23)$$

- 6: **for** index  $1 \leq c \leq i_{ext}$  **do**

- 7: Calculate histogram of left- and right-extension as follow:

$$h_{left} = \sum_{i=mid_x-c}^{mid_x} h_x(i) \quad (4.24)$$

$$h_{right} = \sum_{j=(mid_x-1)+c}^N h_x(j) \quad (4.25)$$

8: Calculate percent of the summation histogram:

$$\text{percent} = (100/\text{sum}) \times (h_{\text{left}} + h_{\text{right}}) \quad (4.26)$$

9: **if**  $\text{percent} < 0.99$  **then**

10: Goto step 6

11: **else**

12:  $i = \text{mid}_x - c$  and  $j = (\text{mid}_x - 1) + c$

13: Go to step 16

14: **end if**

15: **end for**

16: Calculate index of maximum histogram of pixel histogram along with y-direction  $h_y$  to be the middle index of normalized distribution technique.

$$\text{mid}_y = \arg \max_y (h_y(m)) \quad (4.27)$$

17: Calculate summation pixels in the histogram:

$$\text{sum} = \sum_{m=1}^M h_y(m) \quad (4.28)$$

18: Calculate range of extension index

$$i_{\text{ext}} = \begin{cases} N - \text{mid}_y & \text{if } \text{mid}_y \leq N/2 \\ \text{mid}_y & \text{if } \text{mid}_y > N/2 \end{cases} \quad (4.29)$$

19: **for** index  $1 \leq c \leq i_{\text{ext}}$  **do**

20: Calculate histogram of top- and bottom-extension as follow :

$$h_{\text{top}} = \sum_{o=\text{mid}_x-c}^{\text{mid}_x} h_y(i) \quad (4.30)$$

$$h_{\text{bottom}} = \sum_{p=(\text{mid}_y-1)+c}^N h_y(j) \quad (4.31)$$

21: Calculate percent of the summation histogram:

$$\text{percent} = (100/\text{sum}) \times (h_{\text{top}} + h_{\text{bottom}}) \quad (4.32)$$

22: **if**  $\text{percent} < 0.99$  **then**

23: Go to step 19

```

24:  else
25:       $o = mid_y - c$  and  $p = (mid_y - 1) + c$ 
26:      Go to step 29
27:  end if
28: end for
29: Create a new image  $P$  of size  $|i - j| \times |o - p|$  such that  $P(m, n) = I(m, n)$ ;  $i \leq m \leq j$ ;  $o \leq$ 
       $n \leq p$ 

```

#### 4.2.2 Proposed semantic feature extraction

The difference of postures depends upon the curve and angle of player's arm and legs. The posture of a player can be recognized by parts of the body such as how hands raises, how legs bows when hitting or throwing in baseball, how legs bows when kicking or jumping in soccer, or how hands raise when swimming in breaststroke or butterfly styles.

To reach the posture recognition, this study proposes the technique of color highlighting of YCbCr to make the color different between the player region and its background. In YCbCr transformation, only Cb and Cr are selected in the experiment. Moreover, apply the technique of slicing the object into smaller parts and uses 2D Discrete Fourier Transform (2D-DFT) is applied to decompose the shape from the original image of RGB format. From the preliminary experimental, reducing images is improved the accuracy of sport classification. Consequently, reducing image is applied to posture classification.

**Algorithm Blurred background elimination explains** the processes of semantic feature extraction along with Figure 4.6. In Figure 4.5, the input image can be defined into two kinds. First, horizontal image defined by the image of size  $M \times N$ ;  $M < N$ . Second, vertical image defined by the image of size  $M \times N$ ;  $M > N$ .

The semantic feature extraction is process by the following algorithm.

##### **Algorithm Features extraction**

- 1: Eliminate blurred background of image  $I$  of size  $[M \times N]$ .
- 2: Eliminate high-contrast area elimination of image  $R$  of size  $[M \times N]$ .
- 3: Eliminate low-contrast area elimination of image  $H$  of size  $[M \times N]$ .
- 4: Calculate the region of player area selection of image  $L$  of size  $[\dot{M} \times \dot{N}]$ .
- 5: Transform the result image from step 4 to grayscale image denoted by  $I_g$  of size  $[\dot{M} \times \dot{N}]$ .
- 6: **for** scaling factor  $r = 0.2, 0.3, \dots, 0.8$  **do**



- 7: Create a new image  $I'_g$  from  $I_g$  by scaling factor  $r$  and compute the new height  $M'$  and width  $N'$  of  $I_g$ .
- 8: Apply DFT followed by SVD to  $I_g$  and obtain diagonal elements in forms of a vector  $\mathbf{d}_g$ .
- 9: Create a new image  $I'_{Cb}$  from Cb of YCbCr transformation by scaling factor  $r$ .
- 10: **for** number of slices  $2 \leq s \leq 10$  **do**
- 11:     Slice  $I'_{Cb}$  in row-wise fashion such that the height of each slice is equal to  $\lceil M'/s \rceil$  and obtain the set of sliced image denoted by  $U_{Cb}^{(s)}$ .
- 12:     Slice  $I'_{Cb}$  in column-wise fashion such that the height of each slice is equal to  $\lceil N'/s \rceil$  and obtain the set of sliced image denoted by  $V_{Cb}^{(s)}$ .
- 13: **end for**
- 14: Apply DFT followed by SVD to the image in  $U_{Cb}^{(s)}$ , for  $2 \leq s \leq 10$ , and obtain diagonal elements in forms of vectors  $\mathbf{u}_{Cb}^{(s)}$ , for  $2 \leq s \leq 10$ .
- 15: Apply DFT followed by SVD to the image in  $V_{Cb}^{(s)}$ , for  $2 \leq s \leq 10$ , and obtain diagonal elements in forms of vectors  $\mathbf{v}_{Cb}^{(s)}$ , for  $2 \leq s \leq 10$ .
- 16: Create a new image  $I'_{Cr}$  Cr of YCbCr transformation by scaling factor  $r$ .
- 17: **for** number of slices  $2 \leq s \leq 10$  **do**
- 18:     Slice  $I'_{Cr}$  in row-wise fashion such that the height of each slice is equal to  $\lceil M'/s \rceil$  and obtain the set of sliced image denoted by  $U_{Cr}^{(s)}$ .
- 19:     Slice  $I'_{Cr}$  in column-wise fashion such that the height of each slice is equal to  $\lceil N'/s \rceil$  and obtain the set of sliced image denoted by  $V_{Cr}^{(s)}$ .
- 20: **end for**
- 21: Apply DFT followed by SVD to the image in  $U_{Cr}^{(s)}$ , for  $2 \leq s \leq 10$ , and obtain diagonal elements in forms of vectors  $\mathbf{u}_{Cr}^{(s)}$ , for  $2 \leq s \leq 10$ .
- 22: Apply DFT followed by SVD to the image in  $V_{Cr}^{(s)}$ , for  $2 \leq s \leq 10$ , and obtain diagonal elements in forms of vectors  $\mathbf{v}_{Cr}^{(s)}$ , for  $2 \leq s \leq 10$ .
- 23: **end for**
- 24: **for** scaling factor  $r = 0.2, 0.3, \dots, 0.8$  **do**
- 25:     **for** number of slices  $2 \leq s \leq 10$  **do**
- 26:         **if**  $M' > N'$  **then**
- 27:             Concatenate  $\mathbf{u}_{Cb}^{(s)}$ ,  $\mathbf{u}_{Cr}^{(s)}$ ,  $\mathbf{v}_{Cb}^{(s)}$ ,  $\mathbf{v}_{Cr}^{(s)}$ , and  $\mathbf{d}_g$  to form an input vector  $\alpha_{r,s}$  for neural training.
- 28:         **else**
- 29:             Concatenate  $\mathbf{v}_{Cb}^{(s)}$ ,  $\mathbf{v}_{Cr}^{(s)}$ ,  $\mathbf{u}_{Cb}^{(s)}$ ,  $\mathbf{u}_{Cr}^{(s)}$ , and  $\mathbf{d}_g$  to form an input vector  $\alpha_{r,s}$  for neural training.
- 30:         **end if**

31: **end for**

32: **end for**

Setting input feature for comparing in details of resizing and slicing process. Cb and Cr feature extraction based on resizing number from 0.2 to 0.8 (7 sets) and slicing number from 2 to 10 (9 sets); totally  $7 \times 9 = 63$  data sets. An example of dividing image by column and row are presented in the Figures 4.7 and 4.8.

In Figure 4.9(a), from left to right images, the first image is the given image. The second image is the result from blurred background elimination. The third image is the result from irrelevant background elimination. The fourth image is the result from player area selection performed with the fifth image by considering all pixels within the red rectangle. To illustrate the merit of the proposed algorithm, two more examples are also given in Figures 4.9(b) and (c). The resulted sequence of the images in these two figures is the same as that in Figure 4.9(a).



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

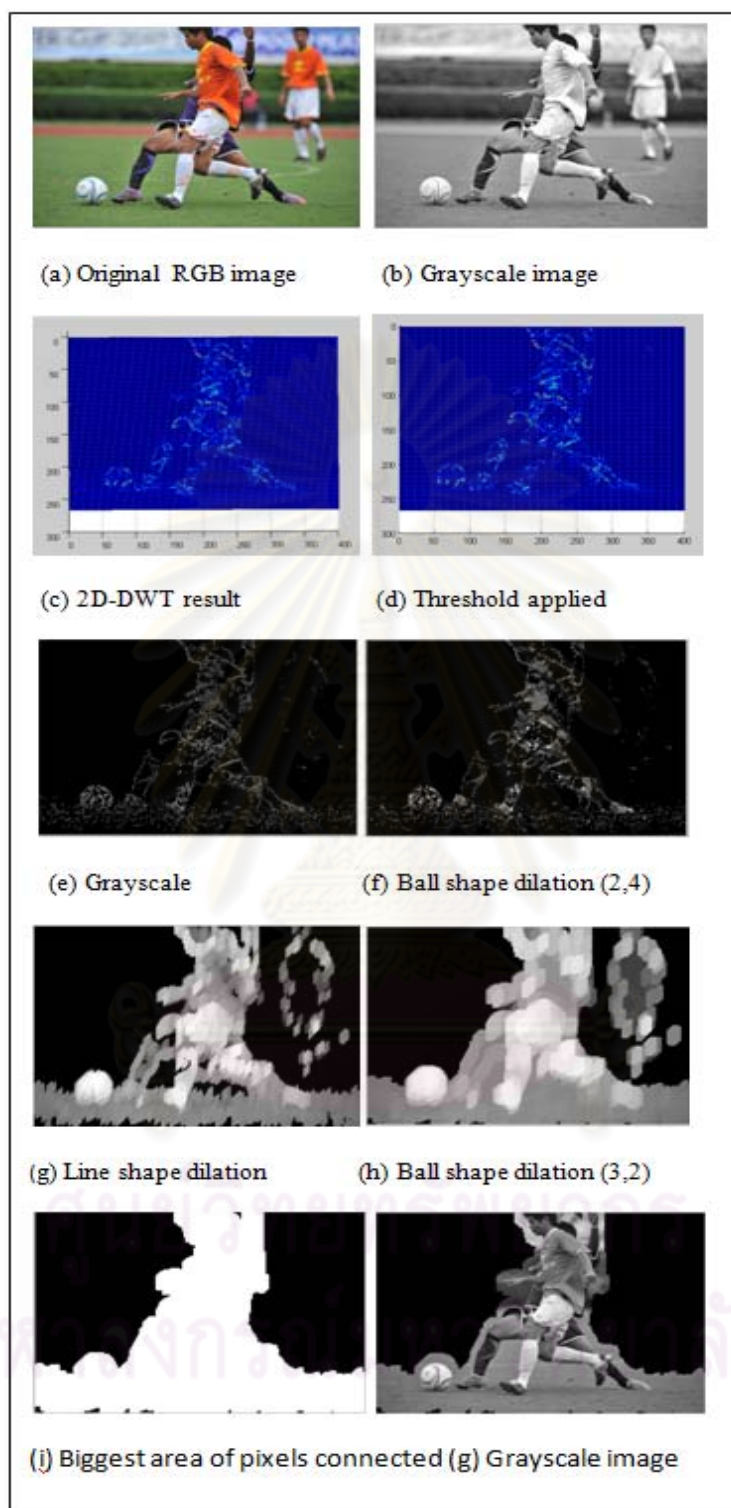


Figure 4.1: Blurred background elimination (a) The original image (b) Value mode of HSV color model. (c) The results of 2D-DWT eliminated (d) Thresholding applied (e) Gray scaled of the result of threshold applied. (f) Dilating by ball shape; radius and height are 2 and 4 pixels. (g) Dilating by line shape kernel. (h) Dilating by ball shape; radius and height are 3 and 2 pixels. (i) Biggest connected pixels. (h) Grayscale image.

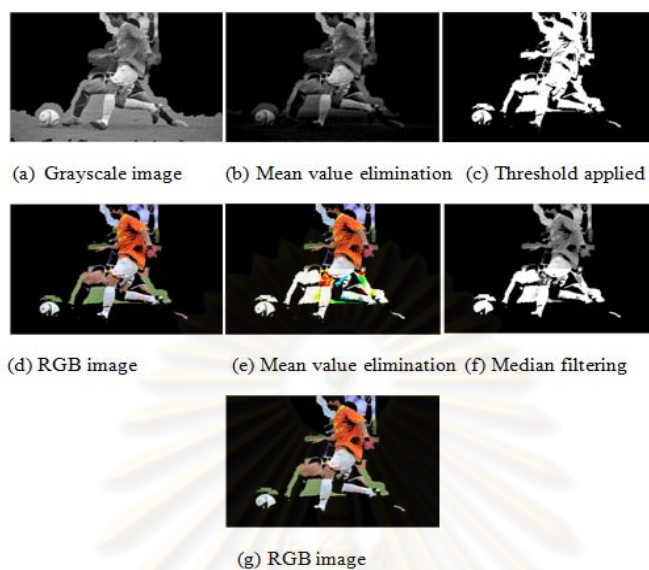


Figure 4.2 High-contrast elimination (top row), and Low-contrast elimination (bottom row).

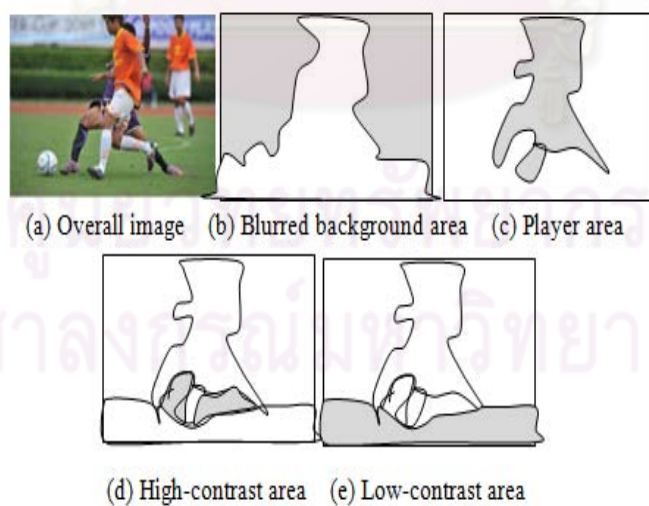


Figure 4.3: Area definition in the highlight area. (a) Original image. (b) Blurred background. (c) Player area. (d) High-contrast area. (e) Low-contrast area.

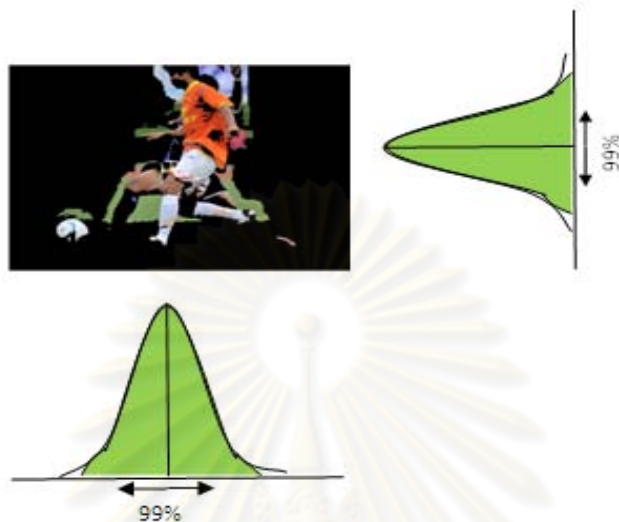


Figure 4.4: The player pixels area has been selected by normal distribution corresponding to pixel histogram in horizontal and vertical direction.

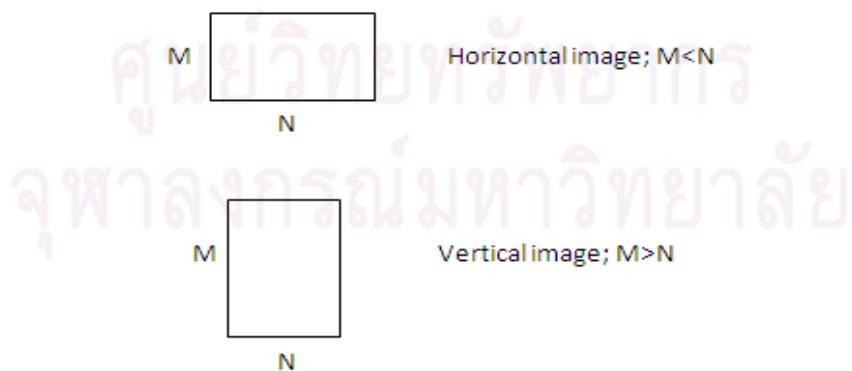


Figure 4.5 Defining horizontal and vertical images.

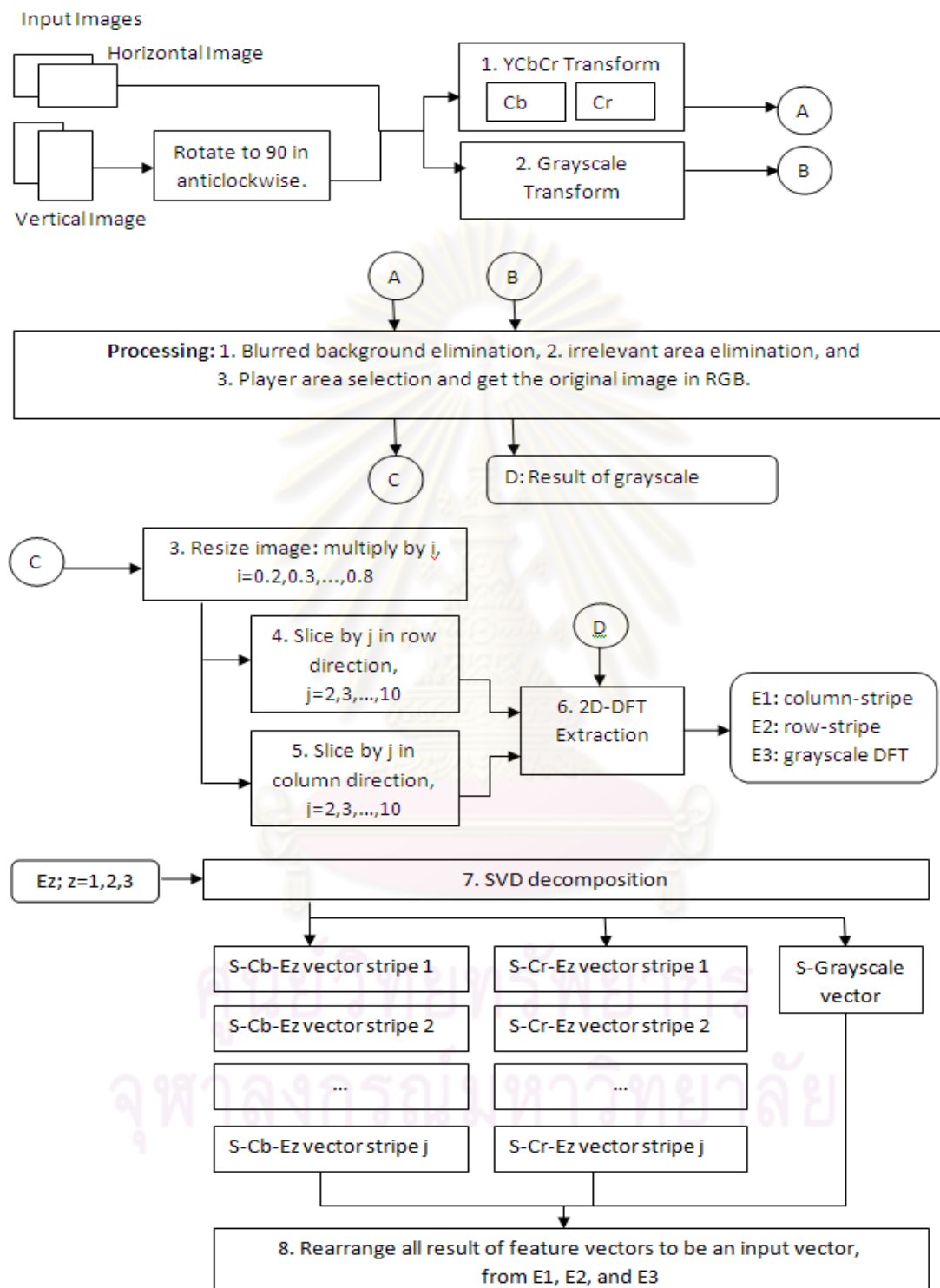


Figure 4.6 Overall feature extraction architecture.



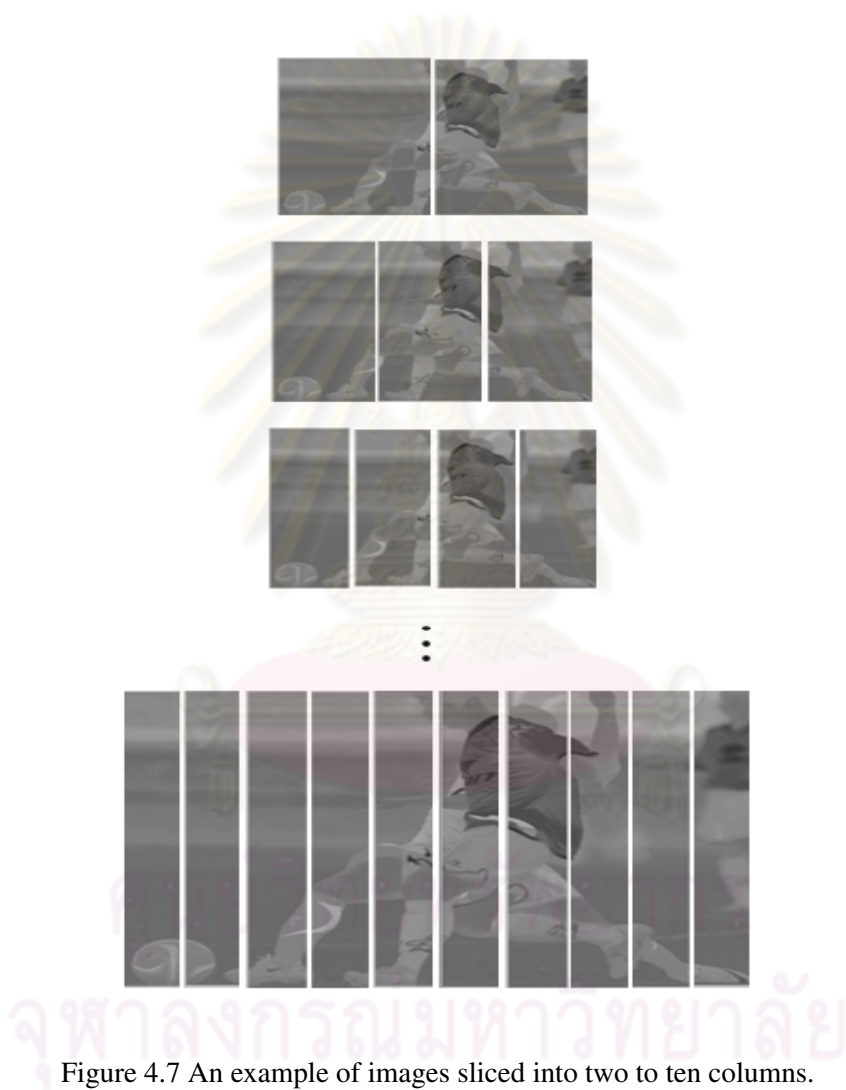




Figure 4.8 An example of images sliced into two to ten rows.



(a)



(b)



(c)

Figure 4.9: Example images of hitting, throwing, and situation postures in baseball from three proposed algorithms.

## CHAPTER V

### EXPERIMENTAL RESULTS

#### 5.1 The overall system

From Figure 5.1 the overall architecture includes two modules 1) sport image classification module, and 2) semantic sport image classification experiments. The experimental results are separated into two phases. First experiment evaluates the effectiveness compared with the other approach, i.e. FNNC. The second experiment evaluates the accuracy of the proposed method by compare proposed feature with EH & RS features.

#### 5.2 Sport image classification

Six sports, namely Baseball, Basketball, Field and Track, Skiing, Soccer, and Swimming are considered in this experiment. There are two directions of experimental comparison. At first, the feature comparison is the experimental comparison between the proposed features and *EH&RS* features proposed in [4]. The first experiment is compared on three neural network architectures, i.e. multi-layer perceptron (MLP), fusion neural network (FNN), and radial basis function (RBF) neural network. In the second direction, our features are compared with the different architectures between NNC and FNNC. Moreover, the result is summarized on the distribution of incorrect classification with the other sports as presented in Tables 5.2-5.3. Totally, 1,800 experimental images are collected from the Internet with 300 images for each sport. In each sport set, 200 images are for training and the other 100 images are for testing. Each image is resized to  $130 \times 200$  pixels. The experimental represented by [47].

In [4], both EH and RS features are parts of *MPEG-7 XM* descriptors. Feature EH is taken from 16 sub-images which are equally partitioned. Then, the spatial distribution in five directional edges, which are horizontal, vertical, two diagonals, and non-directional, are used as the EH features. Hence, there are  $16 \times 5$  features in EH. For RS, 35 dimensions are extracted to capture the distribution of pixels for the shape of object. Combining EH and RS features gives total 115 input features.

For our features, two features sets are used in the experiment. The first set is the decomposition components of the diagonal elements of SVD matrices of DCT, Cb, and Cr matrices. Each matrix

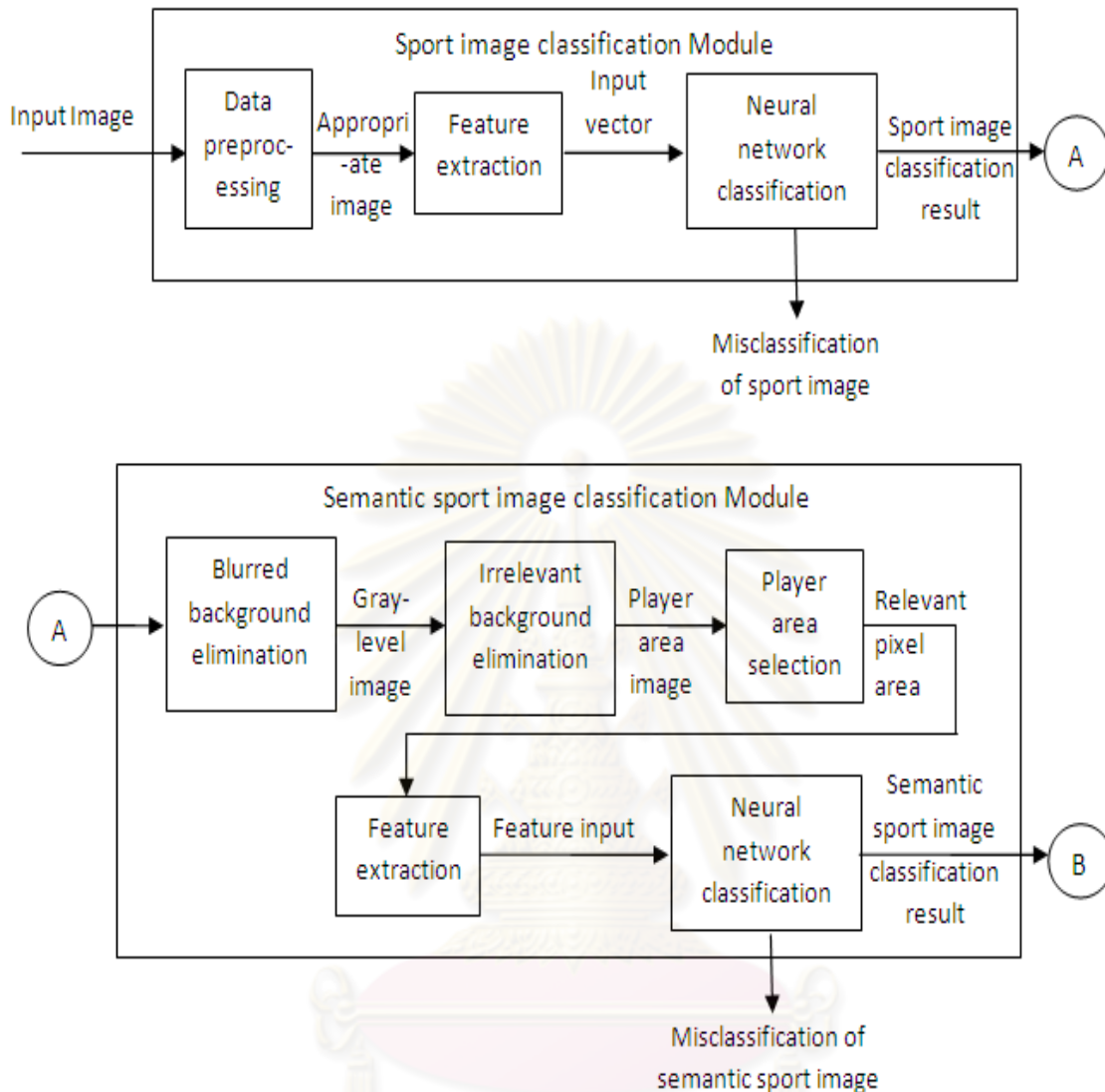


Figure 5.1: The overall process of our proposed, sport image classification module and semantic sport image classification module.

contains a diagonal element as 130 elements. Therefore, the total number of elements from three SVD matrices is  $130 \times 3 = 390$ . The second feature set is from the 64-bin histogram. Totally, there are  $390 + 64 = 454$  input features.

The decimal numbers from the overall input vectors are normalized in the range of  $[0, 1]$ . The architecture of neural network composed of one hidden layer and one output layer. The number of hidden nodes in the hidden layer is 76 hidden nodes and they are trained with the learning rate value of 0.1. For FNN, the weights are adjusted by Levenberg-Marquardt (LM) and resilient propagation

known as *RPROP* method. For MLP and RBF, the weights are adjusted by the technique of RPROP. The reason of using RPROP instead of using LM is due to the memory space. LM uses a larger memory space than RPROP. But LM converges to the goal quicker than RPROP. The *decrease factor*, *increase factor*, and *default factor* are set to 0.5, 1.2, and 0.07, respectively. The activation function in hidden node and output node are sigmoid function and log-sigmoid function. The constraint of stopping error is set to  $10^{-6}$ . Ten trials with new weight initialization in each trail are conducted. The following four sets of epochs 1,000, 5000, 10,000, 20,000 are experimented. All experimented images are carefully selected based on the following criteria.

1. The number of players or people in the images.
  2. Camera focus distance can be short distance, far distance, and in the other view points such as from bottom-to-top direction and top-to-bottom direction depending on the kind of sports.
  3. Texture of background which are ground of sport, audience picture, and billboard background.
- Moreover, the several contrast properties are also considered.

### 5.2.1 Discussion on Features Comparison on NNC, FNN, and RBF

In Table 5.1, our features give higher average accuracy than *EH&RS* features in every network and every number of epochs. For example, in 10000 epoch column, the accuracy of our features versus *EH&RS* features are 82.80% vs 63.45%; 45.17% vs 42.27%; 70.50% vs 54.67% for MLP, FNN, and RBF, respectively. In addition, the standard deviation in our case trends to decrease as the number of epochs increases.

### 5.2.2 Discussion on Distribution of Misclassification

In Table 5.2-5.3, it can be seen that the misclassification of any sport as the other sports is possibly caused by the similarity of players' postures, the textures of player's clothes, or background color. Baseball has the lowest testing accuracy among all sports because of the white or gray color of player's clothes. White or gray color is not outstanding in Cb and Cr color spaces. For example, after 10,000 epochs, Baseball is misclassified as Field&Track, Soccer, Swimming, and Basketball by 15.40%, 12.60%, 3.60%, and 3.40%, respectively. Field & Track has various kinds of sub-sports. The misclassification of this sport is also possibly high. For Soccer, the misclassification may occur because its postures are similar to the postures in many sports such as running or kicking. Among six considered sports, Basketball has the most complex background. So the complex background



Table 5.1: The comparison of classification accuracy between EH & RS and our features on three kinds of neural network.

Epoch		1,000			5,000			10,000			20,000		
		NNC	FNN	RBF	NNC	FNN	RBF	NNC	FNN	RBF	NNC	FNN	RBF
Baseball	EH&RS	45.90	19.80	75.00	49.80	22.50	75.00	53.50	24.90	75.00	58.30	33.40	75.00
	OurFeatures	<b>55.20</b>	<b>34.20</b>	<b>43.00</b>	<b>61.40</b>	<b>18.40</b>	<b>43.00</b>	<b>63.30</b>	<b>19.10</b>	<b>43.00</b>	<b>62.40</b>	<b>38.40</b>	<b>43.00</b>
Basketball	EH&RS	63.50	54.20	62.00	68.30	53.40	62.00	71.20	55.60	62.00	66.70	53.60	62.00
	OurFeatures	<b>77.00</b>	<b>82.70</b>	<b>72.00</b>	<b>83.80</b>	<b>69.80</b>	<b>72.00</b>	<b>85.30</b>	<b>77.60</b>	<b>72.00</b>	<b>86.70</b>	<b>64.50</b>	<b>72.00</b>
Field&Track	EH&RS	63.50	14.20	29.00	67.10	35.90	29.00	60.10	29.80	29.00	61.60	29.60	29.00
	OurFeatures	<b>84.00</b>	<b>9.40</b>	<b>67.00</b>	<b>82.70</b>	<b>43.60</b>	<b>67.00</b>	<b>85.80</b>	<b>37.40</b>	<b>67.00</b>	<b>82.80</b>	<b>47.80</b>	<b>67.00</b>
Skiing	EH&RS	75.00	51.10	52.00	75.60	55.10	52.00	73.20	56.90	52.00	74.60	51.00	52.00
	OurFeatures	<b>88.70</b>	<b>33.40</b>	<b>94.00</b>	<b>91.20</b>	<b>55.10</b>	<b>94.00</b>	<b>90.30</b>	<b>45.40</b>	<b>94.00</b>	<b>90.50</b>	<b>46.70</b>	<b>94.00</b>
Soccer	EH&RS	50.40	7.10	51.00	49.10	36.50	51.00	52.20	36.90	51.00	44.80	40.50	51.00
	OurFeatures	<b>70.90</b>	<b>21.70</b>	<b>65.00</b>	<b>80.30</b>	<b>28.30</b>	<b>65.00</b>	<b>81.10</b>	<b>53.30</b>	<b>65.00</b>	<b>81.50</b>	<b>48.60</b>	<b>65.00</b>
Swimming	EH&RS	62.70	36.40	59.00	74.70	47.10	59.00	70.50	49.50	59.00	72.10	40.50	59.00
	OurFeatures	<b>92.50</b>	<b>19.20</b>	<b>82.00</b>	<b>92.70</b>	<b>41.00</b>	<b>82.00</b>	<b>91.00</b>	<b>38.20</b>	<b>82.00</b>	<b>91.70</b>	<b>56.20</b>	<b>82.00</b>
Mean	EH&RS	60.17	30.47	54.67	64.10	41.75	54.67	63.45	42.27	54.67	63.02	41.43	54.67
	OurFeatures	<b>78.05</b>	<b>33.43</b>	<b>70.50</b>	<b>82.02</b>	<b>42.70</b>	<b>70.50</b>	<b>82.80</b>	<b>45.17</b>	<b>70.50</b>	<b>82.60</b>	<b>50.37</b>	<b>70.50</b>
Std.dev.	EH&RS	9.55	18.02	13.94	10.81	11.36	13.94	8.58	12.45	13.94	9.89	8.62	13.94
	OurFeatures	<b>12.46</b>	<b>23.61</b>	<b>15.73</b>	<b>10.25</b>	<b>16.78</b>	<b>15.73</b>	<b>9.33</b>	<b>17.83</b>	<b>15.73</b>	<b>9.76</b>	<b>8.17</b>	<b>15.73</b>

descriptor of Baseball is not absolutely efficient enough. For Swimming, most of the images show only top part of the swimmer's body and the posture looks similar to the posture of a batter or pitcher in Baseball.

A new set of features for six sports, Baseball, Basket Ball, Swimming, Field & Track, Soccer, and Skiing, is introduced. There are three sets of features representing the majority color of the image, complex background descriptor, and posture descriptor. Regardless of the neural architectures, our features are superior to the previous proposed EH and RS descriptors in terms of classification accuracy. However, these introduced features have not been tested with other kinds of sports. A further study in this issue must be pursued.

### 5.3 Posture classification

The posture classification experiment consists of two systems as shown in Figure 5.1. The input images are classified by the sport image classification at the first filtering. After that, the classified images are classified by the posture classification. The accuracies are evaluated by the performance of classification compare the features between proposed features and EH & RS features.

Since the sport image classification in Chapter III is designed for a horizontal image direction.

Table 5.2: The classification accuracy in each sport compared between NNC and FNN neural network on our features, 1,000 and 5,000 epochs.

		Baseball		Basketball		Field&Track		Skiing		Soccer		Swimming	
		Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
<i>(a) Epoch = 1,000</i>													
<b>Baseball</b>	NNC	<b>75.10</b>	<b>55.20</b>	9.10	6.70	6.00	17.80	1.60	4.10	6.75	8.10	1.45	8.10
	FNN	<b>48.50</b>	<b>34.20</b>	22.40	29.70	4.40	6.80	13.60	13.30	7.05	9.50	4.05	6.50
<b>Basketball</b>	NNC	7.00	3.60	<b>83.45</b>	<b>77.00</b>	4.55	13.90	3.05	2.90	0.60	0.40	1.35	2.20
	FNN	6.55	2.90	<b>78.20</b>	<b>82.70</b>	5.05	3.30	5.40	7.10	2.10	1.80	2.70	2.20
<b>Field&amp;Track</b>	NNC	6.45	2.70	6.00	5.30	<b>80.20</b>	<b>84.00</b>	3.00	3.90	2.15	2.40	2.20	1.70
	FNN	40.40	28.60	24.60	35.70	<b>8.85</b>	<b>9.40</b>	17.10	15.90	2.60	3.80	6.45	6.60
<b>Skiing</b>	NNC	0.65	0.30	3.55	3.50	1.00	1.70	<b>93.05</b>	<b>88.70</b>	0.80	2.80	0.95	3.00
	FNN	37.15	32.00	23.15	23.20	3.55	5.30	<b>31.40</b>	<b>33.40</b>	1.40	1.00	3.35	5.10
<b>Soccer</b>	NNC	11.55	10.60	3.10	1.40	3.05	11.90	3.25	3.70	<b>76.45</b>	<b>70.90</b>	2.60	1.50
	FNN	29.10	17.40	32.25	42.30	2.80	2.60	9.65	10.20	<b>19.80</b>	<b>21.70</b>	6.40	5.80
<b>Swimming</b>	NNC	1.60	2.30	1.10	2.60	0.85	0.80	2.90	1.00	1.60	0.80	<b>91.95</b>	<b>92.50</b>
	FNN	26.40	20.60	27.35	36.20	4.20	3.70	18.65	15.40	3.50	4.90	<b>19.90</b>	<b>19.20</b>
<i>(b) Epoch = 5,000</i>													
<b>Baseball</b>	NNC	<b>89.60</b>	<b>61.40</b>	2.65	3.80	4.75	16.40	0.00	1.10	2.80	12.70	0.20	4.60
	FNN	<b>32.20</b>	<b>18.40</b>	16.15	20.30	17.45	27.40	6.85	9.70	15.85	13.40	11.50	10.80
<b>Basketball</b>	NNC	3.75	1.00	<b>91.00</b>	<b>83.80</b>	3.75	10.90	0.00	0.70	0.70	2.00	0.80	1.60
	FNN	6.30	3.00	<b>71.05</b>	<b>69.80</b>	7.50	13.50	2.90	5.60	5.75	2.30	6.50	5.80
<b>Field&amp;Track</b>	NNC	3.80	4.10	1.95	3.90	<b>91.75</b>	<b>82.70</b>	0.15	3.70	1.80	3.90	0.55	1.70
	FNN	13.70	7.10	11.60	17.30	<b>43.45</b>	<b>43.60</b>	12.00	13.20	8.50	8.60	10.75	10.20
<b>Skiing</b>	NNC	0.00	0.30	0.00	2.10	0.00	1.40	<b>99.95</b>	<b>91.20</b>	0.00	2.70	0.05	2.30
	FNN	7.55	4.90	15.25	14.60	12.30	15.60	<b>53.45</b>	<b>55.10</b>	4.45	4.60	7.00	5.20
<b>Soccer</b>	NNC	3.55	6.50	1.30	1.50	3.10	9.40	0.15	0.90	<b>91.30</b>	<b>80.30</b>	0.60	1.40
	FNN	18.00	11.30	16.50	19.00	12.25	19.30	7.05	9.90	<b>35.50</b>	<b>28.30</b>	10.70	12.20
<b>Swimming</b>	NNC	0.05	1.50	0.00	2.90	0.80	1.20	0.00	0.80	0.25	0.90	<b>98.90</b>	<b>92.70</b>
	FNN	12.05	7.50	8.90	12.20	14.00	16.30	13.75	14.70	8.35	8.30	<b>42.95</b>	<b>41.00</b>

The vertical image is rotated to horizontal image to gain the same aspect ratio. Since the proposed method is rotational-invariant, the images are still correctly classified although the direction of object is change. Due to the features sensitived to image size; all input images are resized to the average size of  $397 \times 594$ . To reduce the input image size, the size of input images must be reduced to the average size by a scale factor of 0.3. The details of sport image classification are as follows:

**Algorithm: The data pre-processing of sport image classification**

1. Rotate the vertical direction image by 90 degree in clockwise direction.
2. Decrease the image size of the input images by multiplying with 0.3 of the original size as  $397 \times 597$ , the reduced size are  $120 \times 180$ . The appropriate size is equal to Phase I size which outperforms for sport kind classification.

Table 5.3: The classification accuracy in each sport compared between NNC and FNN neural network on our features, 10,000 and 20,000 epochs.

		Baseball		Basketball		Field&Track		Skiing		Soccer		Swimming	
		Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
<i>(c) Epoch = 10,000</i>													
<b>Baseball</b>	NNC	<b>93.80</b>	<b>63.30</b>	1.35	3.40	2.55	15.40	0.00	1.70	2.20	12.60	0.10	3.60
	FNN	<b>31.15</b>	<b>19.10</b>	13.15	18.60	28.70	30.40	7.60	6.30	17.00	23.20	2.40	2.40
<b>Basketball</b>	NNC	1.95	0.90	<b>95.30</b>	<b>85.30</b>	2.05	10.10	0.00	0.50	0.15	1.50	0.55	1.70
	FNN	3.75	2.40	<b>73.95</b>	<b>77.60</b>	9.25	8.40	3.30	2.40	4.90	3.50	4.85	5.70
<b>Field&amp;Track</b>	NNC	2.85	2.90	0.90	3.90	<b>94.90</b>	<b>85.80</b>	0.00	2.30	1.25	3.70	0.10	1.40
	FNN	12.25	12.10	14.90	21.50	<b>46.55</b>	<b>37.40</b>	12.10	8.90	9.80	17.20	4.40	2.90
<b>Skiing</b>	NNC	0.00	0.00	0.00	2.60	0.00	1.30	<b>100.00</b>	<b>90.30</b>	0.00	3.20	0.00	2.60
	FNN	10.15	10.80	15.55	14.40	15.30	17.00	<b>44.35</b>	<b>45.40</b>	4.70	3.70	9.95	8.70
<b>Soccer</b>	NNC	2.10	5.80	0.80	1.40	1.25	10.50	0.00	0.40	<b>95.50</b>	<b>81.10</b>	0.35	0.80
	FNN	11.00	7.00	13.15	18.00	12.35	15.10	3.10	4.10	<b>56.60</b>	<b>53.30</b>	3.80	2.50
<b>Swimming</b>	NNC	0.05	2.20	0.00	3.60	0.15	1.40	0.00	0.90	0.30	0.90	<b>99.50</b>	<b>91.00</b>
	FNN	5.90	4.80	11.55	18.90	22.70	20.10	8.60	6.80	13.50	11.20	<b>37.75</b>	<b>38.20</b>
<i>(d) Epoch = 20,000</i>													
<b>Baseball</b>	NNC	<b>93.10</b>	<b>62.40</b>	1.30	3.50	2.50	15.10	0.00	2.00	2.95	12.40	0.15	4.60
	FNN	<b>50.45</b>	<b>38.40</b>	14.10	13.40	13.65	21.80	2.70	3.90	15.70	16.10	3.40	6.40
<b>Basketball</b>	NNC	3.00	1.50	<b>94.00</b>	<b>86.70</b>	2.00	7.90	0.00	0.80	0.50	2.30	0.50	0.80
	FNN	8.45	5.10	<b>66.00</b>	<b>64.50</b>	4.70	12.10	7.20	7.30	7.05	4.00	6.60	7.00
<b>Field&amp;Track</b>	NNC	2.75	3.50	0.30	4.00	<b>95.45</b>	<b>82.80</b>	0.15	2.50	1.35	4.70	0.00	2.50
	FNN	25.75	18.50	11.95	12.00	<b>43.45</b>	<b>47.80</b>	9.15	9.90	4.90	5.20	4.80	6.60
<b>Skiing</b>	NNC	0.00	0.00	0.00	2.70	0.00	1.10	<b>100.00</b>	<b>90.50</b>	0.00	3.10	0.00	2.60
	FNN	18.25	13.90	14.15	13.30	8.50	10.30	<b>48.10</b>	<b>46.70</b>	3.70	2.80	7.30	13.00
<b>Soccer</b>	NNC	2.20	5.50	0.95	2.00	1.20	9.40	0.20	0.80	<b>95.15</b>	<b>81.50</b>	0.30	0.80
	FNN	23.25	16.70	12.70	16.70	5.35	10.50	2.65	2.60	<b>52.25</b>	<b>48.60</b>	3.80	4.90
<b>Swimming</b>	NNC	0.00	2.00	0.00	3.70	0.45	1.40	0.05	0.50	0.10	0.70	<b>99.40</b>	<b>91.70</b>
	FNN	18.85	16.00	6.25	9.90	8.35	9.90	6.50	4.40	3.60	3.60	<b>56.45</b>	<b>56.20</b>

3. *Proposed feature* extraction processes as the same way as Algorithm 1 of Chapter III from steps 2 to 8.

4. *EH & RS features* are extracted from the image size of multiply by 0.3. Finally, the size is of  $120 \times 180$ .

The example of features extraction is illustrated in Figure 5.2. The left picture is the input image and the right image is the four extracted features. Sub-images (a) and (b) are Cb and Cr channel from YCbCr transformation. Sub-image (c) is the 64-bins color histogram. Sub-image (d) is the S-DCT of the gray-level input image.

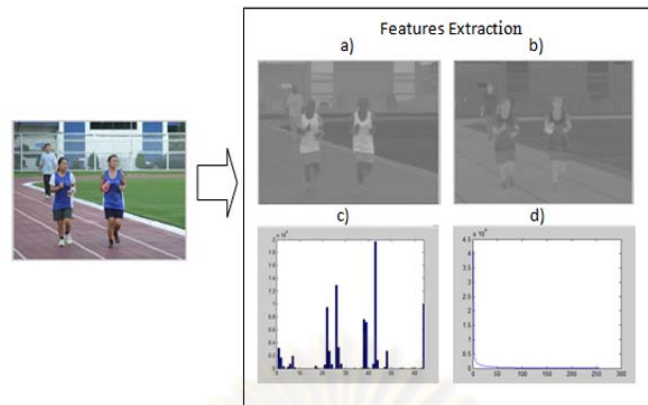


Figure 5.2: The example of feature extraction example for sport image classification a) Cb channel. b) Cr channel. c) 64-bins color histogram. d) S-DCT coefficient.

### 5.3.1 Experimental Set-up for posture classification

The image database was collected from the internet. Both horizontal and vertical directions are considered. There are 2,422 images in the experiment. Because the number is different in each sport which rely on the different quantity of each posture, at least 263 images in basketball and at most 590 images in swimming. The numbers of data in each posture are at least 40 images in field & track and at most 80 images in basketball and skiing. In summary, there are 1,190 images in the experiment. The posture descriptions are presented in Table 5.4.

### 5.3.2 Feature comparison

#### 5.3.2.1 EH & RS features (EH & RS)

*EH* and *RS* features are the elements in *MPEG-7 XM* descriptors. This dissertation used *XM-Tools* for extracting the *EH* and *RS* features. The input vector is formed by concatenating the features along 35 dimensions of *EH*, and 80 dimensions of *RS*, totally 115 dimensions.

#### 5.3.2.2 Proposed features (Pro.Feat.)

This dissertation proposed the feature of 2D-DFT based on resizing and slicing techniques which followed in Chapter IV, Algorithm 1 to 5. The dimensions of each case of resizing and slicing process are different.

Table 5.4 Postures definition.

Sport	Posture Name	Definition
Baseball	Hitting	The batter image with one player.
	Throwing	The throwing image with one player.
	Situation 1	The characteristic on the batter state with two players: the batter and the sitting receiver.
Basketball	Running	Two players of running during the compete game.
	Shooting	Two players of shooting in the encounter game.
	Under goal	The characteristic on taking score scene but the camera viewpoint takes through the ring goal.
Field and Track	Broad Jump	The players of broad jump consists of taking up to the air or taking down to the sand.
	High Jump	The player of high jumping game.
	Group Running	Running of many runners who race in long distance.
	Track Running	Running of sprinting who race in short distance.
	Throwing	One player who is throwing.
Soccer	Jumping	Two players jump to hit a ball with their heads.
	Kicking One	One player kick a ball.
	Running Two	Two player are running to catch a ball.
Skiing	Jumping	The player slice and jump from the upland.
	Turning	The player slice to turn at the corner.
Swimming	Freestyle	The player taking one hand to swim in freestyle.
	Breaststroke	The player taking two hands to swim in breaststroke.
	Butterfly	The player taking two hands to swim in butterfly.

### 5.3.3 Parameter Set-up

Four-fold cross-validation is applied to set up fair training and testing data, i.e. 75% training data and 25% testing data. All input images are classified by two filtering systems, sport image classification system and posture classification system. The parameters are set as follows:

*For sport image classifier: Neural Network set-up*

1. Setting training and testing set with 4-folds cross-validation.
2. Neural network architecture set-up: one hidden layer with 80 hidden nodes, 1,000, 5,000, and 10,000 epochs; 0.1 learning rate; “scaled conjugate gradient” is the training function; “hyperbolic tangent sigmoid” is the transfer function in both of hidden and output layer.
3. Mixing the training and testing sets of each sport from the best accuracy result in step (2).
4. Neural network architecture set-up of the mixed set. Training set relies on three hidden nodes



number of 80, 90, and 100 hidden nodes. Each hidden nodes set is trained in four epochs such as 1,000, 3,000, 5,000 and 10,000 epochs. All of training set of hidden nodes and training's epochs are set-up on six learning rate from 0.1 to 0.6. Training function and transfer function are the same in step (2).

The average accuracy is calculated from ten trials of new initialed weight. To perform the best set of training and testing set, each sport is selected from the best accuracy result from 4-fold data. The result images from sport images classifier are the input images of the posture classification.

For posture image classifier, the related parameters of NNC architecture are defined as 80, 90, and 100 hidden nodes; 300 epochs; 0.4 learning rate; and the other parameters are represented to the above parameters setting. Each sport is trained independently.

#### **5.3.4 Analysis of sport image classification**

Tables 5.5 and 5.6 present the accuracies of the image classification and misclassification of six sports with 0.1 learning rate, 80, 90 and 100 hidden nodes and 1,000, 5,000 and 10,000 epochs. The best accuracies occur when there are 90 hidden nodes and 10,000 epochs are highest in swimming 98.87% and lowest in soccer 80.77%.

From Tables 5.6 at 90 hidden nodes, the most misclassification of six sports occurs in soccer; 14.81% in baseball and 4.42% in field & track and swimming. The reasons of misclassification in postures of hitting or throwing is due to arms and legs stretching which is similar to a soccer player who is running or kicking the ball. For this reason, it is obvious that the most percentage of misclassification occur on soccer and baseball. Also in basketball and field & track, some misclassification also occurs in the general postures. In swimming, the half body of player can be mistaken in the classification.

#### **5.3.5 Analysis of posture recognition**

Tables 5.7 and 5.8 present the experimental results based on 90, 100, and 110 hidden nodes, 300 epochs, and 0.4 learning rate. This table illustrates the posture classification and misclassification in each sport compares between proposed features and EH & RS features. The best accuracy depends on two parameters such as resizing and dividing number. The best parameters in dividing and resizing numbers are in baseball (BB) is 2 and 0.3, basketball (BK) is 2 and 0.3, field & track (FT) is 10 and 0.3, soccer (SC) is 7 and 0.8, skiing (SK) is 10 and 0.3, and swimming (SW) is 7 and 0.5.

From Table 5.9, the best average accuracies are based on 100 hidden nodes. The accuracies of



Table 5.5: The percentage of accuracies in image classification and misclassification of six sports with learning rate 0.1, 1,000 and 5,000 epochs (Notice:ep.=’epochs’, and hd=’hidden nodes’)

1000 ep.	Percent	BB	BK	FT	SC	SK	SW	
80 hd	BB	<b>88.82</b>	1.57	1.96	5.29	1.37	0.98	
	BK	3.77	<b>90.94</b>	2.45	1.32	0.00	1.51	
	FT	11.20	0.40	<b>84.00</b>	3.60	0.80	0.00	
	SC	17.31	0.00	3.27	<b>79.42</b>	0.00	0.00	
	SK	2.11	0.00	2.63	0.53	<b>93.42</b>	1.32	
	SW	0.00	0.00	0.94	0.38	0.00	<b>98.68</b>	
	90 hd	BB	<b>88.43</b>	0.98	1.76	5.88	1.76	1.18
BK	3.77	<b>91.32</b>	2.08	1.51	0.00	1.32		
FT	11.80	0.00	<b>85.40</b>	2.20	0.60	0.00		
SC	19.81	0.00	3.85	<b>76.35</b>	0.00	0.00		
SK	2.89	0.00	2.89	2.63	<b>90.53</b>	1.05		
SW	0.00	0.00	0.75	0.00	0.00	<b>99.25</b>		
100 hd	BB	<b>86.08</b>	1.76	1.76	6.86	1.96	1.57	
	BK	4.15	<b>90.19</b>	2.83	1.51	0.00	1.32	
	FT	9.60	0.40	<b>86.60</b>	3.20	0.20	0.00	
	SC	18.65	0.00	2.50	<b>78.85</b>	0.00	0.00	
	SK	2.37	0.26	3.42	1.58	<b>91.32</b>	1.05	
	SW	0.19	0.19	0.57	0.00	0.00	<b>99.06</b>	
	5000 ep.	Percent	BB	BK	FT	SC	SK	SW
80 hd		BB	<b>86.47</b>	1.76	2.16	6.27	1.96	1.37
		BK	3.58	<b>91.32</b>	2.08	1.70	0.00	1.32
		FT	10.40	0.00	<b>87.40</b>	1.60	0.60	0.00
		SC	17.31	0.38	4.81	<b>77.50</b>	0.00	0.00
		SK	1.58	0.00	2.63	1.58	<b>93.16</b>	1.05
		SW	0.38	0.00	0.00	0.00	0.00	<b>99.62</b>
	90 hd	BB	<b>86.67</b>	1.37	1.57	7.06	1.76	1.57
BK		4.34	<b>90.38</b>	2.08	1.51	0.00	1.70	
FT		10.00	0.20	<b>86.40</b>	3.40	0.00	0.00	
SC		16.15	0.19	3.46	<b>80.19</b>	0.00	0.00	
SK		1.58	0.00	3.16	1.84	<b>92.89</b>	0.53	
SW		0.19	0.00	0.75	0.00	0.00	<b>99.06</b>	
100 hd		BB	<b>88.04</b>	0.98	1.57	6.67	1.57	1.18
	BK	3.77	<b>91.13</b>	2.08	1.89	0.00	1.13	
	FT	10.40	0.00	<b>85.40</b>	3.80	0.40	0.00	
	SC	16.15	0.00	3.65	<b>80.19</b>	0.00	0.00	
	SK	1.05	0.00	2.89	1.32	<b>93.42</b>	1.32	
	SW	0.00	0.19	0.57	0.00	0.00	<b>99.25</b>	

proposed features outperform the accuracies of EH & RS features, especially in baseball, basketball, field & track, soccer, and swimming. Notice that, the proposed feature have the mean accuracy, i.e.

Table 5.6: The percentage of accuracies in image classification and misclassification of six sports with learning rate 0.1, 10,000 epochs. (Notice:ep.=‘epochs’ and hd=‘hidden nodes’)

<i>10000 ep.</i>	Percent	BB	BK	FT	SC	SK	SW
80 hd	BB	<b>87.84</b>	1.18	1.18	6.67	1.96	1.18
	BK	3.96	<b>90.57</b>	2.08	1.32	0.19	1.89
	FT	9.60	0.00	<b>85.80</b>	4.20	0.40	0.00
	SC	16.15	0.00	4.23	<b>79.62</b>	0.00	0.00
	SK	1.32	0.00	2.63	0.79	<b>94.47</b>	0.79
	SW	0.19	0.00	0.38	0.19	0.00	<b>99.25</b>
90 hd	BB	<b>87.84</b>	1.37	1.18	6.86	1.57	1.18
	BK	4.72	<b>90.75</b>	2.26	1.32	0.00	0.94
	FT	8.20	0.00	<b>88.80</b>	3.00	0.00	0.00
	SC	14.81	0.00	4.42	<b>80.77</b>	0.00	0.00
	SK	1.05	0.00	4.21	0.53	<b>92.63</b>	1.58
	SW	0.19	0.19	0.75	0.00	0.00	<b>98.87</b>
100 hd	BB	<b>88.82</b>	0.98	1.57	5.88	1.76	0.98
	BK	4.91	<b>90.75</b>	1.89	0.75	0.00	1.70
	FT	10.40	0.00	<b>86.60</b>	3.00	0.00	0.00
	SC	18.08	0.00	3.46	<b>78.46</b>	0.00	0.00
	SK	2.63	0.00	2.63	1.58	<b>90.79</b>	2.37
	SW	0.00	0.00	1.32	0.00	0.00	<b>98.68</b>

90.86%, is higher than EH & RS feature i.e. 78.35%.

Besides skiing, EH & RS can capture the posture region with the less complex of background. In case of sports having more complex backgrounds, EH & RS reduce the effectiveness of classification while the proposed features do not. For example, basketball has more complex backgrounds because of small area and indoor sport. The proposed features achieves 89.76% but EH & RS achieves 77.76%. Notice that the misclassification occurs on the following three kinds of postures:

- The general posture; running in basketball.
- The posture which has the superimposed of body parts; high jump in field & track.
- The posture which is not unique such as butterfly in swimming.

The reason of low accuracy in soccer is because of the complex background which makes it difficult to detect the region of interest. In case of high jump in field & track, high jump has some region in superimposed parts which difficulty to classify the player shape. Lastly, butterfly posture in swimming is presented only half of the body which is very similar to the other postures of swimming.

*Discussion case 1:* the general posture; running in basketball. From Table 5.7 at the 100 hidden nodes present the accuracies are in running 82.79%, shooting 93.75%, and under goal 92.94%.

Table 5.7: The average accuracies of postures classification and misclassification compare between proposed features(Pro.Feat.) and EH & RS features(EH&RS). (Notice, BB='baseball', BK='basketball', FT='field & track', ep.='epochs', hd='hidden nodes', div.='dividing number', and rez.='resizing number')

		Hitting		Throwing		Situation 1					
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.				
<i>BB div.2/rez.0.3</i>											
80 hd.	Hitting	<b>73.38</b>	<b>97.44</b>	8.13	0.75	18.50	1.81				
	Throwing	1.73	8.13	<b>82.73</b>	<b>91.47</b>	15.53	0.40				
	Situation 1	28.50	3.29	16.64	2.93	<b>54.86</b>	<b>93.79</b>				
90 hd.	Hitting	<b>74.25</b>	<b>98.44</b>	7.63	0.38	18.13	1.19				
	Throwing	1.33	8.73	<b>83.20</b>	<b>91.07</b>	15.47	0.20				
	Situation 1	28.00	3.50	16.93	2.21	<b>55.07</b>	<b>94.29</b>				
100 hd.	Hitting	<b>73.69</b>	<b>98.38</b>	7.94	0.31	18.38	1.31				
	Throwing	2.00	7.93	<b>83.40</b>	<b>91.73</b>	14.60	0.33				
	Situation 1	27.86	3.14	16.86	2.21	<b>55.29</b>	<b>94.64</b>				
<i>BK div.2/rez.0.3</i>											
		Running		Shooting		Under goal					
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.				
80 hd.	Running	<b>69.43</b>	<b>82.14</b>	19.79	2.29	10.79	15.57				
	Shooting	17.94	5.75	<b>74.56</b>	<b>93.94</b>	7.50	0.31				
	Under goal	4.33	1.28	5.50	5.56	<b>90.17</b>	<b>93.17</b>				
90 hd.	Running	<b>70.79</b>	<b>82.79</b>	19.00	1.93	10.21	15.29				
	Shooting	17.56	6.06	<b>75.13</b>	<b>93.75</b>	7.31	0.19				
	Under goal	4.56	1.44	5.39	5.56	<b>90.06</b>	<b>93.00</b>				
100 hd.	Running	<b>68.50</b>	<b>82.57</b>	19.64	2.00	11.86	15.43				
	Shooting	17.50	6.19	<b>74.63</b>	<b>93.75</b>	7.88	0.06				
	Under goal	4.44	1.50	5.39	5.56	<b>90.17</b>	<b>92.94</b>				
<i>FT div.10/rez.0.3</i>											
		Broad Jump		High Jump		Group Running		Track Running		Throwing	
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.
80 hd.	Broad Jump	<b>88.56</b>	<b>99.56</b>	10.89	0.44	0.00	0.00	0.00	0.00	0.56	0.00
	High Jump	18.43	14.29	<b>77.14</b>	<b>82.43</b>	4.43	0.14	0.00	0.00	0.00	3.14
	Group Running	0.13	3.38	10.88	0.38	<b>89.00</b>	<b>93.25</b>	0.00	0.00	0.00	3.00
	Track Running	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	<b>100.00</b>	0.00	0.00
	Throwing	16.63	0.00	0.00	12.50	8.13	0.00	0.00	0.00	<b>75.25</b>	<b>87.50</b>
90 hd.	Broad Jump	<b>88.78</b>	<b>99.56</b>	10.89	0.44	0.11	0.00	0.00	0.00	0.22	0.00
	High Jump	17.43	14.29	<b>78.43</b>	<b>82.71</b>	4.00	0.14	0.00	0.00	0.14	2.86
	Group Running	0.00	3.00	11.00	0.25	<b>89.00</b>	<b>93.25</b>	0.00	0.00	0.00	3.50
	Track Running	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	<b>100.00</b>	0.00	0.00
	Throwing	16.88	0.00	0.13	12.50	8.00	0.00	0.13	0.00	<b>74.88</b>	<b>87.50</b>
100 hd.	Broad Jump	<b>88.78</b>	<b>99.33</b>	10.33	0.67	0.00	0.00	0.00	0.00	0.89	0.00
	High Jump	21.29	14.00	<b>74.71</b>	<b>83.00</b>	3.43	0.00	0.00	0.00	0.57	3.00
	Group Running	0.00	3.13	10.38	0.50	<b>89.63</b>	<b>93.00</b>	0.00	0.00	0.00	3.38
	Track Running	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	<b>100.00</b>	0.00	0.00
	Throwing	17.38	0.00	0.00	12.50	7.38	0.00	0.13	0.00	<b>75.13</b>	<b>87.50</b>

Table 5.8: The average accuracies of postures classification and misclassification. (Notice, SC='soccer', SK='skiing', SW='swimming', ep.='epochs', hd='hidden nodes', div.='dividing number', and rez.='resizing number')

<i>SC div.7/rez.0.8</i>		Jumping		Kicking One		Running Two	
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.
80 hd.	Hitting	<b>66.77</b>	<b>89.38</b>	24.23	2.85	9.00	7.77
	Throwing	6.93	3.21	<b>64.57</b>	<b>89.64</b>	28.50	7.14
	Situation 1	20.36	0.57	19.64	4.50	<b>60.00</b>	<b>94.93</b>
90 hd.	Hitting	<b>66.92</b>	<b>89.23</b>	23.85	2.92	9.23	7.85
	Throwing	6.71	4.36	<b>65.50</b>	<b>88.50</b>	27.79	7.14
	Situation 1	21.29	0.79	18.93	4.71	<b>59.79</b>	<b>94.50</b>
100 hd.	Hitting	<b>66.54</b>	<b>89.54</b>	24.54	2.62	8.92	7.85
	Throwing	7.21	3.21	<b>65.07</b>	<b>89.64</b>	27.71	7.14
	Situation 1	20.00	0.79	20.36	4.43	<b>59.64</b>	<b>94.79</b>

<i>SK div.10/rez.0.3</i>		Jumping		Turning	
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.
80 hd.	Jumping	<b>89.94</b>	<b>88.83</b>	10.06	11.17
	Turning	3.11	9.26	<b>96.89</b>	<b>90.74</b>
90 hd.	Jumping	<b>90.06</b>	<b>88.72</b>	9.94	11.28
	Turning	3.21	9.58	<b>96.79</b>	<b>90.42</b>
100 hd.	Jumping	<b>90.28</b>	<b>88.78</b>	9.72	11.22
	Turning	2.95	9.16	<b>97.05</b>	<b>90.84</b>

<i>SW div.7/rez.0.5</i>		Freestyle		Breaststroke		Butterfly	
		EH&RS	Pro.Feat.	EH&RS	Pro.Feat.	EH&RS	Pro.Feat.
80 hd.	Freestyle	<b>79.78</b>	<b>87.00</b>	4.11	12.94	16.11	0.06
	Breaststroke	24.33	6.11	<b>69.83</b>	<b>92.50</b>	5.83	1.39
	Butterfly	4.81	11.69	8.88	9.50	<b>86.31</b>	<b>78.81</b>
90 hd.	Freestyle	<b>79.33</b>	<b>86.50</b>	4.78	13.50	15.89	0.00
	Breaststroke	24.11	6.67	<b>69.44</b>	<b>92.00</b>	6.44	1.33
	Butterfly	5.31	10.25	8.56	8.56	<b>86.13</b>	<b>81.19</b>
100 hd.	Freestyle	<b>79.83</b>	<b>87.06</b>	4.44	12.94	15.72	0.00
	Breaststroke	23.78	6.11	<b>69.83</b>	<b>92.89</b>	6.39	1.00
	Butterfly	5.38	10.50	8.88	9.00	<b>85.75</b>	<b>80.50</b>

Notice that the misclassification of running occurred in under goal 15.43% and shooting 2.00%. Since background of basketball is high contrast area, the method blurred background elimination cannot eliminate. Consequently, the region of interest is not clear in the real region of player and moreover the 2D-DFT cannot decompose a different coefficient between high contrast information of player area and its background. In conclusion, as well as the background is high contrast, the difficulty of posture capturing is very complicated. An example of misclassification is shown in Figure 5.3 and some examples of postures are presented in Figure 5.4.

*Discussion case 2:* The posture which has the superimposed of body parts; high jump in field &

Table 5.9: The comparison table of average accuracy between proposed features and EH & RS based on 80, 90, and 100 hidden nodes (hd.).

Average accuracy	EH&RS			Pro.Feat.		
	80 hd.	90 hd.	100 hd.	80 hd.	90 hd.	100 hd.
BB	70.32	70.84	70.79	94.23	94.60	94.92
BK	78.05	78.66	77.76	89.75	89.85	89.76
FT	85.99	86.22	85.65	92.55	92.60	92.57
SC	63.78	64.07	63.75	91.32	90.74	91.32
SK	93.42	93.42	93.67	89.79	89.57	89.81
SW	78.64	78.30	78.47	86.10	86.56	86.81
mean	78.37	78.58	<b>78.35</b>	90.62	90.65	<b>90.86</b>

Postures	Running	Shooting	Under Goal
Running			
Shooting			
Under Goal			

Figure 5.3 Example of misclassified image of basketball.

track. From Table 5.7, the accuracy at 100 hidden nodes are broad jump 99.33%, high jump 83.00%, group running 93.00%, track running 100.00%, and throwing 87.50%. Notice, high jump is the minimum accuracy. Misclassification of high jump is occurred in broad jump 14.00% and throwing 3.00%.

For example, area superimposed of player is presented in Figure 5.6 (b) at the middle figure and the last figure. The area superimposed is the head of the player which are presented in the same area. Therefore, the posture is not unique and should be the parts of the other sports, especially similarly to broad jump pose. An example of misclassification is shown in Figure 5.5 and some examples of postures are presented in Figure 5.4.





Figure 5.4 Example of basketball images.



Postures	Broad Jump	High Jump	Group Running	Track Running	Throwing
Broad Jump					
High Jump					
Group Running					
Track Running					
Throwing					

Figure 5.5 Example of misclassified images of field &amp; track.



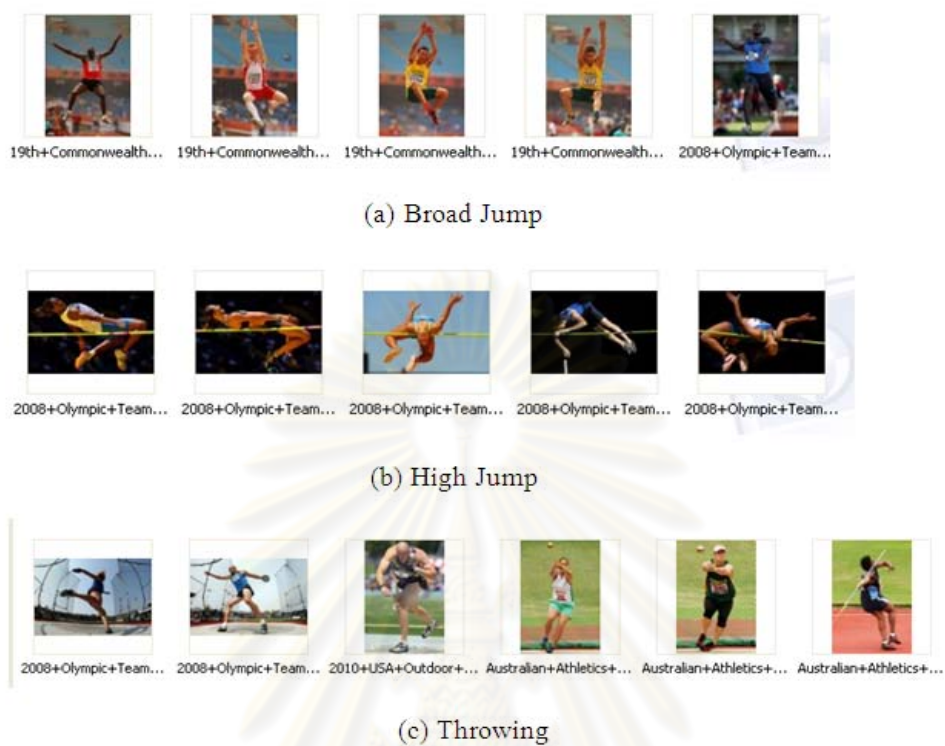


Figure 5.6 Example of field &amp; track images.

Postures	Freestyle	Breaststroke	Butterfly
Freestyle			
Breaststroke			
Butterfly			

Figure 5.7 Example of misclassified images of swimming.

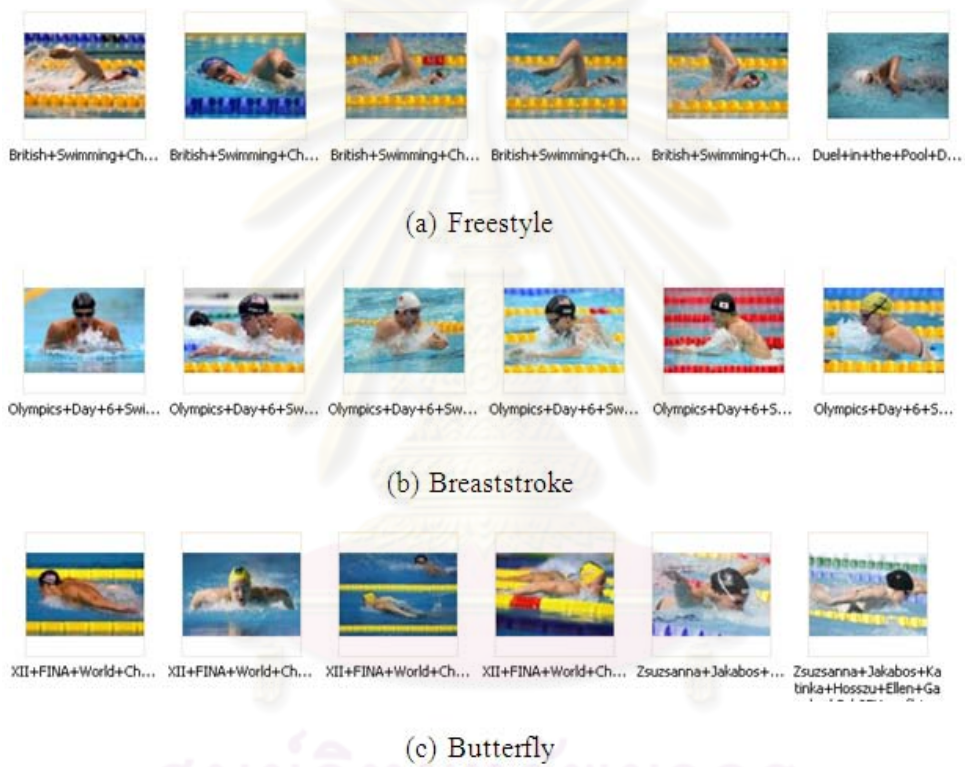


Figure 5.8 Example of swimming images.

## CHAPTER VI

### CONCLUSION

#### 6.1 *Sport classification*

A new set of features for six sports, Baseball, Basketball, Swimming, Field & Track, Soccer, and Skiing, is introduced. There are three sets of features representing the majority color of the image, complex background descriptor, and posture descriptor. Regardless of the neural architectures, our features are superior to previous proposed EH and RS descriptors in terms of classification accuracy. Mean accuracy of the proposed feature is 82.60% and EH & RS features is 63.02%

#### 6.2 *Posture classification*

A new set of algorithms to extract feature for posture classification of nineteen postures is introduced. There are four algorithms for features extraction instead of blurred background elimination, irrelevant area elimination, player area selection, and semantic feature extraction. Two-dimensional Discrete Fourier Transform (2D-DFT) is introduced to be the shape descriptor. Image resizing and slicing are applied to improve the accuracy of posture recognition. The proposed algorithm is robust to object rotation and object size because of 2D-DFT property and player area selection technique.

The proposed method is appropriate for posture changes of body side-view. The proposed algorithm is an outstanding tool for similar postures classification. In similar postures of baseball, the proposed features outperform EH & RS presented at 94.92% and 70.79%. Moreover, the mean accuracy value of the proposed feature is outstanding value as 90.86% while EH & RS features shows at 78.35%.

White color of snow is investigated that be an insensitive color of YCbCr transformation in the color highlighting technique. Consequently, it is the reason of more misclassified image in postures of skiing. Therefore, the technique of color highlighting should be improved. Moreover, misclassified images occurred in two cases are similar postures and superimposed shapes. Local features should be investigated to improve the similar postures classification and overlapping postures classification.

However, these introduced features have not been tested with other kinds of sports. Further study in this issue must be pursued.

## REFERENCES

- [1] Paluri, B., Pradeep, S. N., Shah, H., and Prakash, C., Sports classification using cross-ratio histograms. Proceedings of the 8th Asian conference on Computer vision - Volume Part II, ACCV'07, Berlin, Heidelberg, Springer-Verlag, (2007): 116–123.
- [2] Kamarposhty, M. S., Asadollahi, H., and Teymoori, M. M., Line Detection in Sport Images Using Multi-agent Systems. Proceedings of the 2009 International Conference on Future Computer and Communication, Washington, DC, USA, IEEE Computer Society, (2009): 319–323.
- [3] Jung, Y., Hwang, E., and Kim, W., Sports Image Classification through Bayesian Classifier. CAEPIA, (2003): 546-555.
- [4] Kang, S. and Park, S. A fusion neural network classifier for image classification. Pattern Recognition Letters 30 , 9 (2009): 789 - 793. Advanced Intelligent Computing Theory and Methodology, International Conference on Intelligent Computing.
- [5] Yong-hong, T., Tie-jun, H., and Wen, G. Exploiting multi-context analysis in semantic image classification. Journal of Zhejiang University - Science A 6 (2005): 1268-1283. 10.1007/BF02841665.
- [6] Liu, Y., Zhang, D., and Lu, G. Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recognition 41 , 8 (2008): 2554 - 2570.
- [7] Kesorn, K. and Poslad, S., Enhanced Sports Image Annotation and Retrieval Based Upon Semantic Analysis of Multimodal Cues. Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology, PSIVT '09, Berlin, Heidelberg, Springer-Verlag, (2008): 817–828.
- [8] Rea, N., Dahyot, R., and Kokaram, A., Semantic Event Detection in Sports through Motion Understanding. Proceedings of Conference on Image and Video Retrieval, July 2004.
- [9] Gao, S., Wang, Z., Chia, L.-T., and Tsang, I. W.-H., Automatic image tagging via category label and web data. Proceedings of the international conference on Multimedia, MM '10, New York, NY, USA, ACM, (2010): 1115–1118.

- [10] Li, L.-J. and Fei-Fei, L., What, where and who? Classifying events by scene and object recognition. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (Oct. 2007): 1 -8.
- [11] Bae, T., Kim, C., Jin, S., Kim, K., and Ro, Y., Semantic Event Detection in Structured Video Using Hybrid HMM/SVM. Image and Video Retrieval (Leow, W.-K., Lew, M., Chua, T.-S., Ma, W.-Y., Chaisorn, L., and Bakker, E., eds.) 3568.
- [12] Huang, Y.-P., Chiou, C.-L., and Sandnes, F. E. An intelligent strategy for the automatic detection of highlights in tennis video recordings. Expert Systems with Applications 36 , 6 (2009): 9907 - 9918.
- [13] Tseng, V., Su, J.-H., Ku, H.-H., and Wang, B.-W., Intelligent Concept-Oriented and Content-Based Image Retrieval by using data mining and query decomposition techniques. Multimedia and Expo, 2008 IEEE International Conference on (april 26 2008): 1273 -1276.
- [14] Lin, C. Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network. Pattern Recognition Letters 28 , 16 (2007): 2190 - 2200.
- [15] Demirel, H. and Anbarjafari, G. Pose Invariant Face Recognition Using Probability Distribution Functions in Different Color Channels. Signal Processing Letters, IEEE 15 (2008): 537 -540.
- [16] Zhang, X., Liang, L., Duan, D., and Xia, W., A novel method of face detection based on fusing YCbCr and HIS color space. Automation and Logistics, 2009. ICAL '09. IEEE International Conference on (aug. 2009): 831 -835.
- [17] Ganapathi, T. and Plataniotis, K., Color face recognition under various learning scenarios. Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on (may 2008): 001549 -001552.
- [18] Yang, J. and Liu, C. Horizontal and Vertical 2DPCA-Based Discriminant Analysis for Face Verification on a Large-Scale Database. Information Forensics and Security, IEEE Transactions on 2 (dec. 2007): 781 -792.
- [19] Noda, H., Takao, N., and Niimi, M., Colorization in YCbCr Space and its Application to Improve Quality of JPEG Color Images. Image Processing, 2007. ICIP 2007. IEEE International Conference on 4 (16 2007-oct. 19 2007): IV -385 -IV -388.



- [20] Zhang, C., Feng, X., Li, L., and Song, Y., Identification of cotton contaminants using neighborhood gradient based on YCbCr color space. Signal Processing Systems (ICSPS), 2010 2nd International Conference on 3 (july 2010): V3-733 -V3-738.
- [21] Xie, M., Wu, J., Zhang, L., and Li, C., A novel boiler flame image segmentation and tracking algorithm based on YCbCr color space. Information and Automation, 2009. ICIA '09. International Conference on (june 2009): 138 -143.
- [22] Zhang, X.-N., Jiang, J., Liang, Z.-H., and Liu, C.-L. Skin color enhancement based on favorite skin color in HSV color space. Consumer Electronics, IEEE Transactions on 56 (aug. 2010): 1789 -1793.
- [23] Liu, Z.-M., Li, Y., He, X., Wang, W.-K., Zhou, J.-L., and Li, K., Extraction of face regions in color image. Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on 3 (2002): 1340 - 1343 vol.3.
- [24] Bhatia, A., Srivastava, S., and Agarwal, A., Face Detection Using Fuzzy Logic and Skin Color Segmentation in Images. Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on (nov. 2010): 225 -228.
- [25] Zhang, Z., Gunes, H., and Piccardi, M., An accurate algorithm for head detection based on XYZ and HSV hair and skin color models. Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on (oct. 2008): 1644 -1647.
- [26] Chen, W., Shi, Y., and Xuan, G., Identifying Computer Graphics using HSV Color Model and Statistical Moments of Characteristic Functions. Multimedia and Expo, 2007 IEEE International Conference on (july 2007): 1123 -1126.
- [27] Satonaka, T., Baba, T., Otsuki, T., Chikamura, T., and Meng, T., Object recognition with luminance, rotation and location invariance. Image Processing, 1997. Proceedings., International Conference on 3 (oct 1997): 336 -339 vol.3.
- [28] Huang, W. and Wu, Q., Detection and Tracking of Multiple Moving Objects in Real-World Scenarios using Attributed Relational Graph. Computer and Robot Vision, 2008. CRV '08. Canadian Conference on (may 2008): 245 -252.
- [29] Shermina, J., Illumination invariant face recognition using Discrete Cosine Transform and Principal Component Analysis. Emerging Trends in Electrical and Computer Technology (ICE-TECT), 2011 International Conference on (march 2011): 826 -830.



- [30] Zhu, J. Image compression using wavelets and JPEG2000: a tutorial. Electronics Communication Engineering Journal 14 (jun 2002): 112 -121.
- [31] Zhang, L. and Li, A., Robust Watermarking Scheme Based on Singular Value of Decomposition in DWT Domain. Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on 2 (july 2009): 19 -22.
- [32] Demirel, H., Ozcinar, C., and Anbarjafari, G. Satellite Image Contrast Enhancement Using Discrete Wavelet Transform and Singular Value Decomposition. Geoscience and Remote Sensing Letters, IEEE 7 (april 2010): 333 -337.
- [33] Manian, V. and Vasquez, R., A framework for the recognition of scaled, translated and rotated objects using the short time Fourier transform. Systems, Man, and Cybernetics, 1996., IEEE International Conference on 1 (oct 1996): 730 -735 vol.1.
- [34] Sarfraz, M., Object Recognition using Fourier Descriptors: Some Experiments and Observations. Computer Graphics, Imaging and Visualisation, 2006 International Conference on (july 2006): 281 -286.
- [35] Sarfraz, M., Hassan, M., and Iqbal, M., Object Recognition Using Fourier Descriptors and Genetic Algorithm. Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of (dec. 2009): 318 -323.
- [36] Manshor, N., Rajeswari, M., and Ramachandram, D., Multi-feature Based Object Class Recognition. Digital Image Processing, 2009 International Conference on (march 2009): 324 -329.
- [37] Prasantha, H., Shashidhara, H., and Balasubramanya Murthy, K., Image Compression Using SVD. Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on 3 (dec. 2007): 143 -145.
- [38] Venkateswaran, N., Vignesh, J., Kumar, S., Bharadwaj, M., Rahul, S., and Rao, Y., Hybrid DWT-SVD-VQ Image Compression for Monochrome Images. Signal Processing, Communications and Networking, 2007. ICSCN '07. International Conference on (feb. 2007): 277 -280.
- [39] Dixit, M. and Priyatamkumar, Comparative analysis of variable quantization DCT and variable rank matrix SVD algorithms for image compression applications. Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on (dec. 2010): 1 -5.

- [40] Navas, K., Ajay, M., Lekshmi, M., Archana, T., and Sasikumar, M., DWT-DCT-SVD based watermarking. Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on (Jan. 2008): 271 -274.
- [41] Xie, J., Yang, C., and Huang, D., High Capacity Information Hiding Algorithm for DCT Domain of Image. Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP '08 International Conference on (Aug. 2008): 269 -272.
- [42] Bingabr, M. and Varshney, P. Recovery of corrupted DCT coded images based on reference information. Circuits and Systems for Video Technology, IEEE Transactions on 14 (April 2004): 441 - 449.
- [43] Khayam, S. A. Discrete Cosine Transform (DCT) : Theory and Application.
- [44] Poynton C. A. A Technical Introduction to Digital Video. USA : John Wiley Sons, 1996.
- [45] Noda, H., Takao, N., and Niimi, M., Colorization in YCbCr Space and its Application to Improve Quality of JPEG Color Images. Image Processing, 2007. ICIP 2007. IEEE International Conference on, 16 2007.
- [46] Baker, K., Singular Value Decomposition Tutorial. tech. rep., Ohio State, Mar. 2005.
- [47] Panakarn, P., Phimoltares, S., and Lursinsap, C., Complex sport image classification using spatial color and posture context descriptors and neural classifiers. Machine Learning and Cybernetics (ICMLC), 2010 International Conference on 2 (2010): 713 -718.



Appendix

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## APPENDIX A

### Example of misclassification of posture classification



Figure 1 Example images of soccer are misclassified in baseball.



Figure 2 Example images of basketball are misclassified in baseball.



Figure 3 Example image of field & track is misclassified in baseball.



Figure 4 Example image of skiing is misclassified in baseball.



Figure 5 Example image of baseball is misclassified in basketball.



Figure 6 Example image of baseball is misclassified in field & track.



Figure 7 Example images of basketball are misclassified in field & track.



Figure 8 Example images of soccer are misclassified in field & track.



Figure 9 Example image of skiing is misclassified in field & track.



Figure 10 Example image of swimming is misclassified in field & track.



Figure 11 Example images of baseball are misclassified in soccer.





Figure 12 Example image of basketball is misclassified in soccer.



Figure 13 Example images of field & track are misclassified in soccer.



Figure 14 Example images of baseball are misclassified in skiing.



Figure 15 Example images of baseball are misclassified in swimming.

## Biography

**Name:** Ms. Piyanan Panakarn.

**Date of Birth:** 11<sup>st</sup> September, 1979.

**Educations:**

- Ph.D., Program in Computer Science, Department of Mathematics, Chulalongkorn University, Thailand, (June 2006 - March 2011).
- M.Sc. Program in Information Technology, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, Thailand, (June 2003 - October 2005).
- B.Sc. Program in Computer Science, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, Thailand, (June 1997 - March 2000).

**Publication paper:**

- P. Panakarn, S. Phimoltares, and C. Lursinsap, “Complex sport image classification using spatial color posture context descriptors and neural classifiers,” in *Proc. IEEE ICMLC*, Qingdao, July 2010, pp.713-718.

**Scholarship:**

- The Office of Higher Education Commission, Ministry of Education, Thailand. The scholarship is given pass through Ubon Ratchathani University.
- Science and Technology Innovation Support Grant, Chulalongkorn University