

การจำแนกวิชาทัศนศึกษาประเภทคอร์ทโดยใช้ลักษณะวิชาทัศนระดับล่าง



นายณัฐชัย วัชรากินชัย

สถาบันวิทยบริการ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

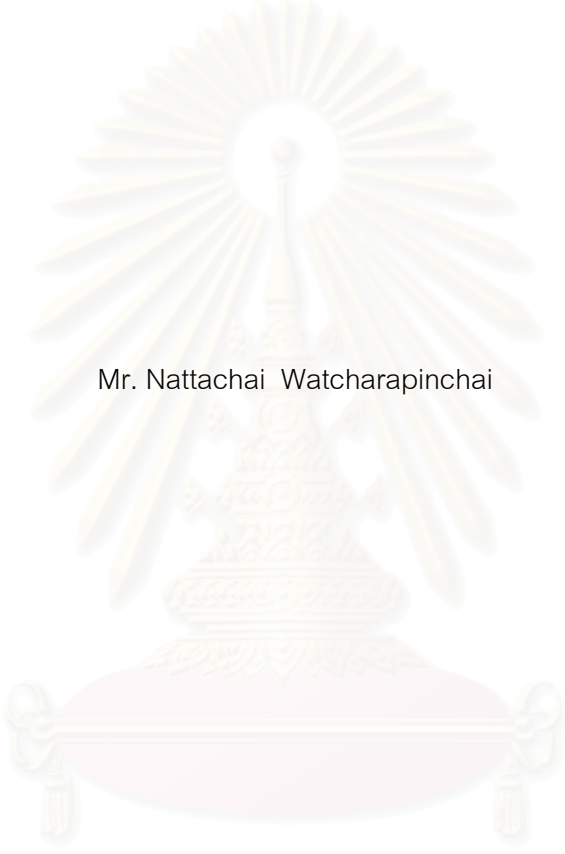
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-53-2845-6

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

COURT SPORT VIDEO CLASSIFICATION USING LOW LEVEL VIDEO FEATURE



Mr. Nattachai Watcharapinchai

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2005


ISBN 974-53-2845-6

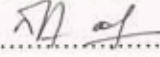
หัวข้อวิทยานิพนธ์	การจำแนกวิถีทัศนิกภาพประเภทคอร์ทโดยใช้ลักษณะวิถีทัศนิกระดับล่าง
โดย	นายณัฐชัย วัชรากินชัย
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.สุภาวดี อร่ามวิทย์
อาจารย์ที่ปรึกษาร่วม	ดร. ศุภกร สิทธิไชย

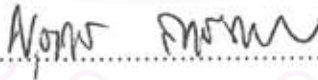
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต



..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร. ดิเรก ลาวณีย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)


..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.สุภาวดี อร่ามวิทย์)


..... อาจารย์ที่ปรึกษาร่วม
(ดร. ศุภกร สิทธิไชย)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เจษฎา ชินรุ่งเรือง)

นายณัฐชัย วัชรากินชัย : การจำแนกวีดิทัศน์กีฬาประเภทคอร์ทโดยใช้ลักษณะวีดิทัศน์ระดับล่าง.
(COURT SPORT VIDEO CLASSIFICATION USING LOW LEVEL VIDEO FEATURE)

อ. ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร. สุภาวดี อร่ามวิทย์, อ.ที่ปรึกษาร่วม : ดร. ศุภกร สิทธิไชย
จำนวนหน้า 135 หน้า. ISBN 974-53-2845-6.

ปัจจุบันเทคโนโลยีทางด้านข้อมูลสื่อประสม เช่น ข้อมูลภาพ และ ข้อมูลวีดิทัศน์ มีการใช้งานอย่างแพร่หลาย ส่งผลให้จำนวนข้อมูลสื่อประสมมีจำนวนมากขึ้นอย่างรวดเร็ว ดังนั้นจึงมีความจำเป็นต้องมีการจัดเก็บ และ ทำดัชนีวีดิทัศน์ที่มีประสิทธิภาพในการค้นคืน ตัวอย่างการประยุกต์การใช้งานหนึ่งที่ต้องการทำดัชนีวีดิทัศน์ คือ สถานีการแพร่สัญญาณโทรทัศน์ ซึ่งประกอบด้วยวีดิทัศน์หลายประเภท เช่น วีดิทัศน์ประเภทข่าว วีดิทัศน์โฆษณา และ วีดิทัศน์ประเภทกีฬา เป็นต้น ทั้งนี้วีดิทัศน์กีฬาเป็นวีดิทัศน์ที่ผู้ชมโดยทั่วไปให้ความสนใจและได้รับความนิยมจากอดีตถึงปัจจุบัน เช่น ฟุตบอล บาสเกตบอล เทนนิส และ มวย เป็นต้น ซึ่งตัวอย่างกีฬาเหล่านี้ในสถานีการแพร่สัญญาณโทรทัศน์จะมีเพิ่มจำนวนมากขึ้นเรื่อยๆ จึงจำเป็นต้องมีการพัฒนาเทคนิคการจำแนกวีดิทัศน์อัตโนมัติซึ่งมีความสำคัญต่อระบบการจัดเก็บและค้นคืนวีดิทัศน์ ดังนั้นในวิทยานิพนธ์นี้ จึงได้ศึกษาและวิเคราะห์ประสิทธิภาพในการจำแนกประเภทวีดิทัศน์กีฬาด้วย ลักษณะวีดิทัศน์ระดับล่าง ประกอบด้วย ฮิสโตแกรมสี ออโตคอร์รีโลแกรม และ ขอบสี และได้นำเสนอเทคนิคการจำแนกวีดิทัศน์กีฬาด้วยการนำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิเวรอลเพื่อปรับปรุงความแม่นยำของการจำแนก จากผลการทดลองพบว่าออโตคอร์รีโลแกรมเป็นลักษณะวีดิทัศน์ระดับล่างที่ให้ผลความแม่นยำในการจำแนกประเภทกีฬามากที่สุด และ พบว่าการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิเวรอลแบบมีชั้นซ่อน 1 ชั้น เป็นตัวจำแนกที่สามารถเพิ่มความแม่นยำในการจำแนกวีดิทัศน์กีฬาได้ ผลการทดลองจำแนกกีฬา 7 ประเภทได้แก่ บาสเกตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด เทนนิส และ วอลเลย์บอล โดยวิธีการดังกล่าวนี้ให้ผลความแม่นยำที่ 92.5% โดยเทคนิคการจำแนกที่นำเสนอดังกล่าวสามารถนำมาใช้กับการทำดัชนีวีดิทัศน์กีฬา และ การใช้งานการให้คำประกอบวีดิทัศน์กีฬา ต่อไป

ภาควิชาวิศวกรรมไฟฟ้า.....
สาขาวิชาวิศวกรรมไฟฟ้า.....
ปีการศึกษา2548.....

ลายมือชื่อนิสิต.....ณัฐชัย วัชรากินชัย.....
ลายมือชื่ออาจารย์ที่ปรึกษา.....ศุภกร สิทธิไชย.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....สุภาวดี อร่ามวิทย์.....

#4670297221 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: VIDEO CLASSIFICATION / VIDEO ANOTATION / VIDEO INDEXING

NATTACHAI WATCHARAPINCHAI : COURT SPORT VIDEO CLASSIFICATION
USING LOW LEVEL VIDEO FEATURE. THESIS ADVISOR : ASST. PROF
SUPAVADEE ARAMVITH, THESIS COADVISOR : SUPAKORN SIDDHICHAI, Ph. D. ,
135 pp. ISBN 974-53-2845-6.

Presently, the advances in multimedia information such as image and video rapidly progress. As a result, there are an increasing number of multimedia materials including several types of video materials such as news videos, commercial videos, and sports videos. Thus, automatic video indexing and classification are desirable steps for accessing and storing contents in any large video library. Most widely televised sports videos include football, basketball, and tennis, among others. In this thesis, we have investigated and analyzed the performance of low level video features, including color histogram, autocorrelogram, camera motion, and color edge. A principal component analysis within multilayer neural network is proposed to improve the classification accuracy. Experimental results suggest that autocorrelogram can be used to efficiently represent sports video genres. As it offers better classification performance than other low level features. Moreover, a hidden layer neural network with principal components can improve classification performance of neural network. Classification of 7 types of sports videos including basketball, Thai boxing, football, golf, springboard, tennis and volleyball resulted in 92.5% in classification accuracy. The proposed technique can also be extended necessary step toward the sport video indexing and sports video annotation.

Department.....Electrical Engineering.....Student's signature.....*Nattachai Watcharapinchai*
Field of study.....Electrical Engineering.....Advisor's signature.....*Synda Awong*
Academic year2005..... Co-advisor's signature.....*Supadee Aramvith*

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลืออย่างดียิ่งของท่านอาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ. ดร. สุภาวดี อร่ามวิทย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร. ศุภกร สิริธิไชย และ ผู้ให้คำปรึกษาในการวิจัยท่านอื่นๆ ได้แก่ ดร. สรรพฤทธิ์ มฤคทัต และ นักวิจัย NECTEC ทุกท่าน ซึ่งได้รับความกรุณาให้คำปรึกษาและเสนอแนะในข้อคิดเห็นต่าง ๆ ในการทำวิจัยมาโดยตลอด ผู้ทำการวิจัยจึงขอกราบขอบพระคุณ

ขอขอบคุณ สถาบันจิตวิทยาศาสตร์และเทคโนโลยีไทย สำนักพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่ให้ความสนับสนุนในส่วนของคุณทุนสำหรับการวิจัย ซึ่งได้ให้งานวิจัยเป็นไปได้อย่างราบรื่น รวมทั้งคุณอุไร คุณวิภาดา และ คุณจันทร์ธิดา ผู้ซึ่งให้ความช่วยเหลือเป็นอย่างดีในการติดต่อประสานงานเรื่องราวต่าง กับทางสถาบัน

ขอกราบขอบพระคุณ สุรจิตร์ วัชรภินชัย (บิดา) คติยา วัชรภินชัย (มารดา) และครอบครัว ผู้ซึ่งคอยให้กำลังใจ ความรักและความห่วงใยแก่ผู้วิจัยจึงทำให้วิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี

ขอบคุณเพื่อน ๆ พี่ๆ และ น้องๆ ภายในห้องปฏิบัติการกรรมวิธีประมวลสัญญาณดิจิทัล สำหรับกำลังใจ ความช่วยเหลือ คำแนะนำ และ สิ่งดี ๆ ที่มีคุณค่าทางจิตใจ จึงทำให้วิทยานิพนธ์นี้สำเร็จเสร็จสิ้นในที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฎ
สารบัญภาพ.....	ฏ
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 งานวิจัยที่เกี่ยวข้อง.....	2
1.3 ระบบการจำแนกประเภทวิถีทัศนกีฬาโดยใช้ลักษณะเฉพาะระดับล่างของวิถีทัศน.....	5
1.4 เทคนิคที่น่าเสนอ.....	6
1.5 วัตถุประสงค์ของการวิจัย.....	6
1.6 ขอบเขตของการวิจัย.....	6
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
1.8 วิธีดำเนินการวิจัย.....	7
2. ความรู้พื้นฐาน.....	8
2.1 การแยกส่วนวิถีทัศนทางเวลา.....	8
2.1.1 ประเภทของการเปลี่ยนแปลงฉาก.....	9
2.1.2 เทคนิคการแยกส่วนวิถีทัศนทางเวลา.....	11
2.1.3 การเลือกขีดเริ่มเปลี่ยนเพื่อตรวจหารอยต่อของขีด.....	13
2.2 การดึงเฟรมสำคัญ.....	15
2.2.1 แนวความคิดแบบรอยต่อของขีด.....	16
2.2.2 แนวความคิดแบบเนื้อหาด้วยการมองเห็น.....	16
2.2.3 แนวความคิดแบบการวิเคราะห์การเคลื่อนที่.....	17
2.2.4 แนวความคิดแบบการจัดกลุ่ม.....	18
2.3 ฉากวิถีทัศนกีฬา.....	19
2.4 โครงสร้างของระบบการจำแนกวิถีทัศนกีฬาด้วยลักษณะเฉพาะวิถีทัศน.....	20

2.4.1 การจำแนกข้อตเบื้องต้น	20
2.4.2 การจำแนกลักษณะเฉพาะวัตถุกราฟฟิก.....	20
2.4.3 การจำแนกลักษณะเฉพาะของการมองเห็น	20
2.5 ลักษณะระดับล่างของวิถีทัศน์.....	21
2.5.1 ลักษณะสี.....	21
2.5.2 ลักษณะการเคลื่อนที่ของกล้อง.....	26
2.5.3 ลักษณะขอบ	30
2.6 การจำแนกประเภทวิถีทัศน์โดยใช้การรู้จำรูปแบบ.....	36
2.6.1 โครงข่ายนิเวรอล.....	38
2.6.2 การวิเคราะห์ส่วนประกอบमुखสำคัญ.....	42
2.6.3 การวิเคราะห์ดีสคริมีแนนต์เชิงเส้นบนปริภูมิผู้ศูนย์	45
3. เทคนิคที่นำเสนอ	51
3.1 การวิเคราะห์ความถูกต้องของการจำแนกประเภทวิถีทัศน์กีฬาโดยใช้ลักษณะระดับล่าง ของวิถีทัศน์กีฬา.....	52
3.1.1 ลักษณะเฉพาะสี.....	52
3.1.2 ลักษณะเฉพาะการเคลื่อนที่ของกล้อง.....	55
3.1.3 ลักษณะเฉพาะขอบสี	55
3.2 การปรับปรุงความถูกต้องของการจำแนกประเภทวิถีทัศน์กีฬาโดยใช้ตัวจำแนกโครงข่าย นิเวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ.....	56
3.2.1 การวิเคราะห์ส่วนประกอบमुखสำคัญ.....	57
3.2.2 ตัวจำแนกโครงข่ายนิเวรอล	58
3.3 การปรับปรุงความถูกต้องของการจำแนกประเภทวิถีทัศน์กีฬาโดยใช้ตัวจำแนกปริภูมิผู้ ศูนย์.....	59
4. ผลการทดลอง	61
4.1 การทดลองที่ 1 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิถีทัศน์กีฬาของ ลักษณะเฉพาะฮิสโตแกรมสี โดยใช้โครงข่ายนิเวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुख สำคัญ.....	61

4.2 การทดลองที่ 2 วิเคราะห์ความแม่นยำของการจำแนกประเภทวีดิทัศน์กีฬาของ
ลักษณะเฉพาะขอโตคอรีโรแกรมโดยใช้โครงข่ายนิรอรลร่วมกับการวิเคราะห์
ส่วนประกอบมุขสำคัญ67

4.3 การทดลองที่ 3 วิเคราะห์ความแม่นยำของการจำแนกประเภทวีดิทัศน์กีฬาของ
ลักษณะเฉพาะการเคลื่อนที่ของกล็องโดยใช้โครงข่ายนิรอรลร่วมกับการวิเคราะห์
ส่วนประกอบมุขสำคัญ73

4.4 การทดลองที่ 4 วิเคราะห์ความแม่นยำของการจำแนกประเภทวีดิทัศน์กีฬาของ
ลักษณะเฉพาะขอบสีโดยใช้โครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ77

4.5 การทดลองที่ 5 วิเคราะห์ความแม่นยำของการจำแนกประเภทวีดิทัศน์กีฬาเมื่อจำนวน
ประเภทกีฬามีจำนวนมากขึ้นโดยใช้โครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบมุข
สำคัญ.....83

 4.5.1 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้ฮิสโตแกรมสีโดย
 โครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ85

 4.5.2 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้อโตคอรีโรแกรม
 โดยโครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ90

 4.5.3 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้การเคลื่อนที่ของกล็อง
 โดยโครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ94

 4.5.4 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยขอบสีโดยโครงข่ายนิรอรล
 ร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ98

4.6 การทดลองที่ 6 วิเคราะห์ความแม่นยำของการจำแนกประเภทวีดิทัศน์กีฬาด้วยตัว
จำแนกปริภูมิผู้ศูนย์..... 103

 4.6.1 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้ฮิสโตแกรมสีโดยตัว
 จำแนกในปริภูมิผู้ศูนย์..... 104

 4.6.2 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้อโตคอรีโรแกรม
 โดยตัวจำแนกในปริภูมิผู้ศูนย์..... 108

 4.6.3 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้การเคลื่อนที่ของกล็อง
 โดยตัวจำแนกในปริภูมิผู้ศูนย์..... 112

 4.6.4 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้ขอบสีโดยตัวจำแนก
 ในปริภูมิผู้ศูนย์ 116

บทที่	หน้า
4.7 สรุปผลการทดลอง	120
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	122
5.1 สรุปผลการวิจัย.....	122
5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต.....	123
รายการอ้างอิง.....	124
ภาคผนวก.....	129
ประวัติผู้เขียนวิทยานิพนธ์.....	135



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตาราง	หน้า
ตารางที่ 4.1 ความแม่นยำของการจำแนกด้วยโครงข่ายแบบไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะฮิสโตแกรมสี.....	63
ตารางที่ 4.2 คอนฟิวชันเมตริกซ์ของฮิสโตแกรมสีที่ให้ผลความแม่นยำที่ดีที่สุด	66
ตารางที่ 4.3 คอนฟิวชันเมตริกซ์ของฮิสโตแกรมสีที่ให้ผลความแม่นยำที่ดีที่สุด	72
ตารางที่ 4.4 ความแม่นยำของการจำแนกด้วยโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล้อง.....	74
ตารางที่ 4.5 คอนฟิวชันเมตริกซ์ของการเคลื่อนที่ของกล้องที่ให้ผลความแม่นยำที่ดีที่สุด.....	77
ตารางที่ 4.6 ความแม่นยำของการจำแนกด้วยโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสี	79
ตารางที่ 4.7 คอนฟิวชันเมตริกซ์ของขอบสีที่ให้ผลความแม่นยำที่ดีที่สุด	82
ตารางที่ 4.8 จำนวนข้อมูลฝึกฝนและข้อมูลทดสอบ สำหรับ 7 ประเภทกีฬา.....	85
ตารางที่ 4.9 คอนฟิวชันเมตริกซ์ของฮิสโตแกรมสำหรับ 7 ประเภทกีฬา	89
ตารางที่ 4.10 คอนฟิวชันเมตริกซ์ของฮิสโตแกรมสำหรับ 7 ประเภทกีฬา.....	94
ตารางที่ 4.11 คอนฟิวชันเมตริกซ์ของการเคลื่อนที่ของกล้องสำหรับ 7 ประเภทกีฬา.....	98
ตารางที่ 4.12 คอนฟิวชันเมตริกซ์ของขอบสีสำหรับ 7 ประเภทกีฬา	102
ตารางที่ 4.13 คอนฟิวชันเมตริกซ์ของปริภูมิสี่คูณ สำหรับฮิสโตแกรม 7 ประเภทกีฬา	108
ตารางที่ 4.14 คอนฟิวชันเมตริกซ์ของปริภูมิสี่คูณ สำหรับฮิสโตแกรม 7 ประเภทกีฬา .	112
ตารางที่ 4.15 คอนฟิวชันเมตริกซ์ของปริภูมิสี่คูณ สำหรับการเคลื่อนที่กล้อง 7 ประเภทกีฬา..	116
ตารางที่ 4.16 คอนฟิวชันเมตริกซ์ของปริภูมิสี่คูณ สำหรับขอบสี 7 ประเภทกีฬา	120

สารบัญภาพ

ภาพประกอบ	หน้า
รูปที่ 2.1 โครงสร้างระดับวีดิทัศน์.....	9
รูปที่ 2.2 รอยต่อของข้อต่อแบบการเปลี่ยนแปลงทันที	10
รูปที่ 2.3 การเปลี่ยนแปลงฉากแบบเฟดออก.....	10
รูปที่ 2.4 การเปลี่ยนแปลงฉากแบบเฟดขึ้น	10
รูปที่ 2.5 การเปลี่ยนแปลงฉากแบบปาด	10
รูปที่ 2.6 การหาขีดเริ่มเปลี่ยนใจแบบรวม (T) และ แบบปรับตัวได้ ($T(k)$) [5].....	14
รูปที่ 2.7 วิธีการหาขีดเริ่มเปลี่ยนใจโดยใช้หน้าต่างเลื่อน [5]	15
รูปที่ 2.8 การตั้งเฟรมสำคัญโดยระเบียบวิธีการของ Wolf [25]	18
รูปที่ 2.9 สหสัมพันธ์เชิงพื้นที่ระหว่างจุดภาพ P_1 และ P_2	22
รูปที่ 2.10 ตัวอย่างการพิจารณาจุดภาพ (x_i, y_i) ที่ระยะ k เท่ากับ 1	24
รูปที่ 2.11 ตัวอย่างการพิจารณาจุดภาพ (x_i, y_i) ที่ระยะ k เท่ากับ 2	24
รูปที่ 2.12 ภาพตัวอย่างที่มีฮิสโตแกรมของสีเหมือนกัน.....	26
รูปที่ 2.13 ฮิสโตแกรมของภาพตัวอย่างในรูปที่ 2.12	26
รูปที่ 2.14 ลักษณะการเคลื่อนที่ของวัตถุแบบต่าง	28
รูปที่ 2.15 การกำหนดทิศทางหลักของเวกเตอร์การเคลื่อนที่	30
รูปที่ 2.16 ค่าของจุดภาพ	33
รูปที่ 2.17 ตัวพรางสำหรับแนวแกน x	33
รูปที่ 2.18 ตัวพรางสำหรับแนวแกน y	33
รูปที่ 2.19 แผนภูมิสายงานการตรวจหาขอบสี	34
รูปที่ 2.20 แผนภาพการวิเคราะห์แนวคิดเชิงความหมาย	36
รูปที่ 2.21 โครงข่ายนิรวลหลายชั้นแบบไปข้างหน้า	39
รูปที่ 2.22 การแปลงข้อมูลตัวอย่างลงบนส่วนประกอบमुखสำคัญใน 2 มิติ	44
รูปที่ 2.23 การทำให้เมทริกซ์ทแยงมุมพร้อมกัน.....	49
รูปที่ 3.1 ภาพตัวอย่าง.....	53
รูปที่ 3.2 ฮิสโตแกรมของภาพขนาด 8x8 จุดภาพ	53
รูปที่ 3.3 ฮิสโตแกรมของภาพขนาด 200x100 จุดภาพ	54

ภาพประกอบ	หน้า
รูปที่ 3.4 แผนภาพขั้นตอนการจำแนกประเภทวีดิทัศน์โดยใช้โครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ.....	57
รูปที่ 3.5 แผนภาพขั้นตอนการจำแนกประเภทวีดิทัศน์โดยใช้ตัวจำแนกปริภูมิผู้ศูนย์.....	59
รูปที่ 4.1 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับลักษณะเฉพาะฮิสโตแกรมสี.....	63
รูปที่ 4.2 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะฮิสโตแกรมสี	64
รูปที่ 4.3 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน.....	64
รูปที่ 4.4 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ	65
รูปที่ 4.5 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วมของข้อมูลทดสอบ สำหรับ ลักษณะเฉพาะฮอโตคอรีโลแกรม	68
รูปที่ 4.6 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับ ฮอโตคอรีโลแกรม	69
รูปที่ 4.7 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮอโตคอรีโลแกรม	70
รูปที่ 4.8 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮอโตคอรีโลแกรมของข้อมูลฝึกฝน.....	70
รูปที่ 4.9 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮอโตคอรีโลแกรมของข้อมูลทดสอบ	71
รูปที่ 4.10 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล้อง	75
รูปที่ 4.11 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล้อง	75
รูปที่ 4.12 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล้องของข้อมูลฝึกฝน.....	76
รูปที่ 4.13 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล้องของข้อมูลทดสอบ	76

ภาพประกอบ	หน้า
รูปที่ 4.28 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับ การเคลื่อนที่ของกล่องของข้อมูลฝึกฝน.....	95
รูปที่ 4.29 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การใช้การเคลื่อนที่ของกล่องของข้อมูลทดสอบ	95
รูปที่ 4.30 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล่อง.....	96
รูปที่ 4.31 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล่องของข้อมูลฝึกฝน	96
รูปที่ 4.32 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ การเคลื่อนที่ของกล่องของข้อมูลทดสอบ.....	97
รูปที่ 4.33 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน.....	99
รูปที่ 4.34 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับ ขอบสีของข้อมูลทดสอบ	99
รูปที่ 4.35 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสี.....	100
รูปที่ 4.36 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน	100
รูปที่ 4.37 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน	101
รูปที่ 4.38 ผลการทดลองตัวจำแนกปริภูมิผู้ศุนย์ สำหรับ ฮิสโตแกรมสีจากข้อมูลฝึกฝน.....	105
รูปที่ 4.39 ผลการทดลองตัวจำแนกปริภูมิผู้ศุนย์ สำหรับ ฮิสโตแกรมสีจากข้อมูลทดสอบ	105
รูปที่ 4.40 ขนาดปริภูมิผู้ศุนย์ของฮิสโตแกรมสี.....	106
รูปที่ 4.41 การเปรียบเทียบตัวจำแนก สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน.....	106
รูปที่ 4.42 การเปรียบเทียบตัวจำแนก สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ	107
รูปที่ 4.43 ผลการทดลองตัวจำแนกปริภูมิผู้ศุนย์ สำหรับ ออกโตคอร์รีโลแกรมจากข้อมูลฝึกฝน....	109

ภาพประกอบ	หน้า
รูปที่ 4.44 ผลการทดลองตัวจำแนกปริภูมิผู้ศูนย์ สำหรับ ออโตคอร์รีโลแกรมจากข้อมูลทดสอบ .	109
รูปที่ 4.45 ขนาดปริภูมิผู้ศูนย์ของออโตคอร์รีโลแกรม	110
รูปที่ 4.46 การเปรียบเทียบตัวจำแนก สำหรับ ออโตคอร์รีโลแกรมของข้อมูลฝึกฝน.....	110
รูปที่ 4.47 การเปรียบเทียบตัวจำแนก สำหรับ ออโตคอร์รีโลแกรมของข้อมูลทดสอบ	111
รูปที่ 4.48 ผลการทดลองตัวจำแนกปริภูมิผู้ศูนย์ สำหรับ การเคลื่อนที่ของกล่องจากข้อมูลฝึกฝน	113
รูปที่ 4.49 ผลการทดลองตัวจำแนกปริภูมิผู้ศูนย์ สำหรับ การเคลื่อนที่ของกล่องจากข้อมูลทดสอบ.....	113
รูปที่ 4.50 ขนาดปริภูมิผู้ศูนย์ของการเคลื่อนที่ของกล่อง.....	114
รูปที่ 4.51 การเปรียบเทียบตัวจำแนกโดยใช้การเคลื่อนที่ของกล่องของข้อมูลฝึกฝน.....	114
รูปที่ 4.52 การเปรียบเทียบตัวจำแนกโดยใช้การเคลื่อนที่ของกล่องของข้อมูลทดสอบ	115
รูปที่ 4.53 ผลการทดลองตัวจำแนกปริภูมิผู้ศูนย์ สำหรับ ขอบสีจากข้อมูลฝึกฝน.....	117
รูปที่ 4.54 ผลการทดลองตัวจำแนกปริภูมิผู้ศูนย์ สำหรับ ขอบสีจากข้อมูลทดสอบ	117
รูปที่ 4.55 ขนาดปริภูมิผู้ศูนย์ของขอบสี	118
รูปที่ 4.56 การเปรียบเทียบตัวจำแนก สำหรับ ขอบสีของข้อมูลฝึกฝน.....	118
รูปที่ 4.57 การเปรียบเทียบตัวจำแนก สำหรับ ขอบสีของข้อมูลทดสอบ.....	119

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันเทคโนโลยีทางด้านข้อมูลสื่อประสม (multimedia data) เช่น ข้อมูลภาพ ข้อมูลวีดิทัศน์ มีการใช้งานอย่างแพร่หลาย ส่งผลให้จำนวนข้อมูลสื่อประสมมีจำนวนมากขึ้น ดังนั้นจึงมีความจำเป็นต้องมีการจัดเก็บ และ ทำดัชนีวีดิทัศน์ที่มีประสิทธิภาพในการค้นคืน เพื่อเป็นประโยชน์ในการสอบถามจากผู้ใช้งานฐานข้อมูลสื่อประสม

การแพร่สัญญาณโทรทัศน์ (Broadcast Television) เป็นตัวอย่างการประยุกต์การใช้งานข้อมูลวีดิทัศน์แบบหนึ่ง วีดิทัศน์สามารถแบ่งได้หลายประเภท ได้แก่ วีดิทัศน์ประเภทข่าว วีดิทัศน์โฆษณา และ วีดิทัศน์กีฬา เป็นต้น ทั้งนี้วีดิทัศน์กีฬาเป็นวีดิทัศน์ประเภทหนึ่งที่ถูกชมโดยทั่วไปให้ความสนใจมากที่สุด ตัวอย่างวีดิทัศน์กีฬาที่ได้รับความนิยม ได้แก่ ฟุตบอล บาสเก็ตบอล เทนนิส และ มวย เป็นต้น ซึ่งทางสถานีโทรทัศน์แพร่สัญญาณวีดิทัศน์กีฬาในอดีตถึงปัจจุบันการจัดเก็บวีดิทัศน์กีฬาแบบแอนะล็อกสามารถทำได้โดยเก็บข้อมูลเหล่านี้ลงม้วนเทป ซึ่งขาดประสิทธิภาพในแง่ที่ใช้ในการจัดเก็บ ความเสื่อมของเนื้อเทปเมื่อจัดเก็บไว้นาน การเก็บข้อมูลดังกล่าวมาติดต่อเป็นไปด้วยความยุ่งยาก รวมถึงการค้นคืนส่วนข้อมูลวีดิทัศน์อาจต้องใช้เวลานาน จึงปฏิเสธไม่ได้ว่าข้อมูลเหล่านี้มีปริมาณมากและมีมูลค่ามหาศาลกับทางสถานีโทรทัศน์ ดังนั้นจึงจำเป็นต้องมีระบบการจัดเก็บและจัดการข้อมูลวีดิทัศน์ที่มีประสิทธิภาพทั้งในแง่ของการจัดเก็บและการค้นคืน ซึ่งเทคนิคการแยกส่วนภาพและจำแนกประเภทวีดิทัศน์ให้ตรงกับความเข้าใจของมนุษย์เป็นสิ่งจำเป็น ทางเรื่องหนึ่งในปัจจุบันคือการใช้แรงคนในการแยกส่วนภาพและจำแนกวีดิทัศน์ซึ่งใช้เวลามาก และหากจำนวนวีดิทัศน์ที่มีมากขึ้น ความน่าเชื่อถือในการจำแนกจะขึ้นอยู่กับคนที่ทำการจำแนกและปัจจัยทางกายภาพของมนุษย์ ดังนั้นการพัฒนาเทคนิคการจำแนกวีดิทัศน์อัตโนมัติจึงมีความสำคัญต่อระบบการจัดเก็บและค้นคืนวีดิทัศน์ เทคนิคของจำแนกวีดิทัศน์ที่ถูกนำมาใช้ส่วนใหญ่ ได้แก่ การหาความสัมพันธ์ของเวกเตอร์ลักษณะเฉพาะที่ตรงกับลักษณะเชิงความหมาย

ในการพิจารณาข้อมูลวีดิทัศน์ซึ่งประกอบด้วย วัตถุ (object) เหตุการณ์ (event) และฉาก (scene) การอธิบายความสัมพันธ์ระหว่างองค์ประกอบเหล่านี้ เรียกว่า แนวความคิดเชิง

ความหมาย (semantic concept) ตัวอย่างเช่น ผู้สอบถามต้องการสอบถามวิถีทัศน์ที่แสดงถึง “การทำแถมในการแข่งขันบาสเก็ตบอล” ระบบค้นคืนจะต้องหาวิถีทัศน์ที่เกี่ยวกับกีฬาบาสเก็ตบอลซึ่งถือว่าเป็นฉาก ส่วนเหตุการณ์และวัตถุคือการทำแถมและตัวผู้เล่นตามลำดับ ลักษณะนี้เรียกว่าลักษณะระดับสูง (high-level feature) อย่างไรก็ตามลักษณะระดับสูงนี้ไม่สามารถถูกอธิบายโดยลักษณะระดับต่ำ (low-level feature) เช่น สี (color) รูปร่าง (shape) พื้นผิว (texture) ขอบ (edge) หรือ เวกเตอร์การเคลื่อนที่ (motion vector) เพียงเวกเตอร์ใดเวกเตอร์หนึ่งได้โดยง่าย ดังนั้นปัญหาของแนวความคิดเชิงความหมายนั้นก็คือการหาความสัมพันธ์ระหว่างลักษณะระดับต่ำกับลักษณะระดับสูงซึ่งถือว่าเป็นช่องว่างระหว่างลักษณะระดับสูงและลักษณะระดับต่ำที่เรียกว่า ช่องว่างทางความหมาย (semantic gap) ระบบการจำแนกที่ดีควรที่จะมีช่องว่างทางความหมายน้อยที่สุด เพราะลักษณะระดับต่ำสามารถหาได้โดยใช้เทคนิคการประมวลผลภาพ ในขณะที่ลักษณะระดับสูงเป็นความหมายตามความเข้าใจของมนุษย์

เนื่องจากเหตุผลดังกล่าวนี้จึงได้มีผู้นำเสนองานวิจัย [1-3] เกี่ยวกับเวกเตอร์ลักษณะเชิงความหมาย (Semantic feature) หรือ อาจเรียกว่า เวกเตอร์ลักษณะระดับสูง (High Level Feature Vector) ขึ้น โดยที่นำเวกเตอร์ลักษณะระดับต่ำมาสร้างแบบจำลองตรวจหาลักษณะเชิงแนวความหมาย (Semantic feature) หรือ สร้างตัวจำแนก เพื่อนำมาอธิบายวัตถุเหตุการณ์ หรือ ฉากที่เกิดขึ้นในวิถีทัศน์ เพื่อใช้ในการทำดัชนีสำหรับการค้นคืนวิถีทัศน์และการให้คำประกอบวิถีทัศน์ (Video Annotation)

1.2 งานวิจัยที่เกี่ยวข้อง

ในการพัฒนาระบบจัดเก็บอย่างอัตโนมัติจะเริ่มที่ขั้นตอนของการจำแนกวิถีทัศน์เบื้องต้นให้ตรงตามชนิดกีฬาแล้วจึงนำไปสร้างลักษณะเฉพาะเชิงความหมายหรือเหตุการณ์ในแต่ละแบบจำลองกีฬา ซึ่งจะเป็นประโยชน์ต่อผู้ใช้งานที่ต้องการค้นคืนวิถีทัศน์ตามประเภทกีฬาของกีฬา ปัญหาการจำแนกวิถีทัศน์นี้เป็นงานวิจัยที่ได้รับความสนใจอย่างแพร่หลาย ดังจะนำเสนอต่อไปนี้

Gibert และ คณะ [4] นำเสนอเทคนิคการจำแนกประเภทวิถีทัศน์โดยใช้แบบจำลอง HMMs (Hidden Markov Models) ให้กับแต่ละประเภทกีฬา โดยการสร้างแบบจำลอง HMMs จำนวน 2 แบบจำลองของแต่ละประเภทกีฬาจากลำดับลักษณะเฉพาะทิศทางมุมสำคัญสี (principal color direction) และ ลำดับลักษณะเฉพาะทิศทางมุมสำคัญการเคลื่อนที่หลัก (principal motion direction) ที่ได้จากแต่ละเฟรม โดยพิจารณาประเภทของกีฬาได้จากแบบจำลองที่สร้างจากลักษณะเฉพาะทั้ง 2 แบบเป็นอิสระจากกันซึ่งเป็นการคูณกันของความ

น่าจะเป็นที่ได้จากในแต่ละแบบจำลองแล้วเลือกประเภทของกีฬาที่ให้ค่าความน่าจะเป็นมากที่สุด เพื่อจำแนกประเภทกีฬา 4 ประเภท ได้แก่ iceshockey (ice hockey) basketball (basketball) football (football) และ soccer (soccer) ซึ่งผลความแม่นยำที่ได้ประมาณร้อยละ 93

Wang และ คณะ [5] ได้นำเสนอการจำแนกข้อดีวีทัศน์ข่าวกีฬา โดยใช้การตัดสินใจแบบต้นไม้ C4.5 (C4.5 Decision Tree) จากลักษณะเฉพาะของอัตราส่วนสีที่เป็นสนามภายในเฟรมซึ่งอัตราส่วนของสีที่นำมาพิจารณา ได้แก่ อัตราส่วนของสีเหลือง อัตราส่วนของสีเขียว และ อัตราส่วนของสีขาว นอกจากนี้ Wang ได้นำลักษณะเฉพาะของการเคลื่อนไหว และ อัตราส่วนของฉากหลักที่เคลื่อนไหว รวมเป็น 11 เวกเตอร์ลักษณะเฉพาะในการจำแนกข้อดีวีทัศน์ข่าวกีฬา เพื่อจำแนกประเภทกีฬา 4 ประเภท ได้แก่ basketball เบสบอล (baseball) iceshockey และ golf (golf) ซึ่งผลความแม่นยำเป็นร้อยละ 88

Xu และ คณะ [6] ได้สร้างตัวจำแนกโดยใช้แบบจำลอง GMMs (Gaussian Mixture Models) จากลักษณะเฉพาะของการมองเห็น และ ลักษณะเฉพาะของเสียงในช่วงเวลาหนึ่ง และ ลดขนาดของปริภูมิลักษณะเฉพาะทั้งหมดด้วยการวิเคราะห์ส่วนประกอบสำคัญ (principal component analysis) เป็นเวกเตอร์ลักษณะเฉพาะ เพื่อจำแนกประเภทวีดิทัศน์จากสถานีโทรทัศน์ 5 ประเภท ได้แก่ วีดิทัศน์กีฬา วีดิทัศน์การ์ตูน วีดิทัศน์ข่าว วีดิทัศน์โฆษณา และ วีดิทัศน์เพลง ซึ่งผลความแม่นยำเป็นร้อยละ 86.5

Assfalg และ คณะ [2] ใช้โครงข่ายนิเวศในการจำแนกโดยใช้ลักษณะเฉพาะของฮิสโตแกรมสีในปริภูมิสี HSV และ ฮิสโตแกรมของความเข้มขอบ (edge intensity histogram) ซึ่งบทความนี้บ่งบอกว่าถ้าภาพเหตุการณ์ในการจำแนกเป็นเหตุการณ์การแข่งขันกีฬาที่เกิดขึ้นเป็นพื้นที่ที่มีการแข่งขันเท่านั้นจะให้ผลการจำแนกที่ดีกว่าเมื่อมีภาพในเหตุการณ์อื่นๆ ประปน เช่น ภาพเหตุการณ์ผู้ชม ภาพเหตุการณ์แสดงตัวผู้เล่น ซึ่งได้ผลร้อยละ 76.95 โดยงานวิจัยในกลุ่มนี้มุ่งเน้นการให้ค่าประกอบวีดิทัศน์กีฬาดังในบทความ [12 -15] โดยที่มุ่งการวิจัยไปทั้งด้านการเลือกเวกเตอร์ลักษณะระดับล่างที่เหมาะสมและการสร้างตัวจำแนก และ กฎการตัดสินใจที่เหมาะสมกับการให้ค่าประกอบวีดิทัศน์ประเภทกีฬา

Ma และ Zhang [7] นำ SVMs (Support Vector Machines) มาจำแนกโดยใช้ลักษณะเฉพาะของการเคลื่อนไหวของพื้นผิว (Motion Texture Feature) โดย SVMs เป็นวิธีที่ได้จากการประยุกต์ทฤษฎีพื้นฐานทางสถิติส่งผลการจำแนกให้ผลที่ดีในหลากหลายการประยุกต์การใช้งาน แต่ไม่สามารถควบคุมจำนวนเวกเตอร์สนับสนุนได้ บางครั้งแบบจำลองในการจำแนกจึงมี

ขนาดใหญ่ทำให้ใช้เวลาในการประมวลผลของการจำแนกเป็นเวลามาก ซึ่งความแม่นยำที่ได้เป็นร้อยละ 93.57

S. Takagi และ คณะ [10] นำเสนอการวิเคราะห์การจัดกลุ่มให้กับวิดีโอที่บันทึกภาพที่แพร่ภาพจากสถานีโทรทัศน์ โดยใช้พารามิเตอร์ของการเคลื่อนที่ของกล้อง ซึ่งวิธีที่นำเสนอเรียกว่า อัตราส่วนการดึงลักษณะการเคลื่อนที่ของกล้อง (camera motion extraction ratio) และการเปลี่ยนของการเคลื่อนที่ของกล้อง (camera motion transtion) และใช้หลักการวิเคราะห์ทางสถิติในแต่ละกีฬาเพื่อจำแนกประเภทกีฬาซึ่งค่าทางสถิติในการวิเคราะห์ได้แก่ ค่าเฉลี่ย (mean) ค่าความแปรปรวน (variance) ความเบ้ (skewness) และ เคอร์โทซิส (kurtosis)

Messer และ คณะ [8] นำลักษณะเฉพาะระดับล่างสร้างคิวเชิงความหมาย (semantic cue) หลายชนิด ได้แก่ คิวหญ้า คิวลูสนามว่ายนํ้า คิวกลุ่มกองเชียร์ เป็นต้น แล้วจึงนำไปจำแนกชนิดกีฬาโดยอาศัยตัวจำแนกแบบ k-NN ซึ่งผลความแม่นยำเป็นร้อยละ 93.7 นอกจากนี้ Messer และ คณะ [11] นำเสนอวิธีการตัดสินใจแบบแผนภูมิต้นไม้เพื่อจัดหมวดหมู่ให้กับข้อต่อโดยใช้ข้อมูลจากคิวเชิงความหมายที่ได้ [8] และ ใช้การจัดหมวดหมู่โดยใช้ HMMs อีกครั้งหนึ่ง

Sadlier และ Connor [9] สร้างตัวตรวจหา (detector) เพื่อเป็นลักษณะระดับสูงของเหตุการณ์สำหรับทุกประเภทกีฬา ซึ่งตัวตรวจหาที่สร้างขึ้นได้แก่ ตัวตรวจหาภาพฝูงชนบนอัฒจันทร์ ตัวตรวจหากิจกรรมบนช่วงความถี่ของเสียงพูด ตัวตรวจหาวัตถุกราฟฟิกบนภาพ ตัวตรวจหากิจกรรมการเคลื่อนที่ภายในภาพ และ ตัวตรวจหาตำแหน่งของเส้นขอบสนาม โดยที่ตัวตรวจหาเหล่านี้ถือเป็นลักษณะระดับสูง เมื่อนำตัวตรวจหาเหล่านี้มาประกอบรวมกัน เรียกว่า เวกเตอร์ลักษณะเฉพาะของข้อต่อ (shot feature vectors) แล้วนำมาเป็นอินพุตให้กับตัวจำแนก SVM เพื่อสร้างแบบจำลองให้เหตุการณ์ในแต่ละประเภทกีฬาต่อไป

Duan และ คณะ [16] นำเสนอตัวแทนข้อมูลในระดับกลาง (mid-level representation) โดยใช้ความรู้เบื้องต้นเพื่อสร้างเวกเตอร์ลักษณะเฉพาะในระดับกลาง เช่น Motion Entropy, Camera motion parameter เป็นต้น ก่อนที่จะนำมาสร้างลักษณะเฉพาะระดับสูงเช่น เหตุการณ์ต่อไป ซึ่งความแม่นยำที่ได้เป็นร้อยละ 85

ระบบการจำแนกประเภทวิดีโอที่บันทึกภาพส่วนใหญ่ ลักษณะเฉพาะวิดีโอที่ระดับล่างส่วนใหญ่ที่นำไปใช้ คือ ลักษณะเฉพาะระดับล่างของสี และ ลักษณะเฉพาะของการเคลื่อนที่ ลักษณะเฉพาะของสีเป็นลักษณะเฉพาะพื้นฐานที่สามารถนำมาแบ่งแยกประเภทวิดีโอที่บันทึกภาพได้

เนื่องจากในหลากหลายประเภทกีฬาสามารถแยกแยะได้ด้วยลักษณะของสี ตัวอย่างเช่นสีของสนามในแต่ละประเภทกีฬามีความแตกต่างกัน เป็นต้น สำหรับลักษณะเฉพาะของการเคลื่อนที่เป็นลักษณะเฉพาะอีกประเภทหนึ่งที่น่าสนใจ เนื่องจากในกีฬาบางประเภทมีการเคลื่อนที่แตกต่างจากกีฬาประเภทอื่นๆ ทำให้ง่ายต่อการจำแนก ดังนั้นวิทยานิพนธ์นี้จึงทำการค้นหาลักษณะวิถีทัศนระดับล่างที่เหมาะสมแต่การจำแนกประเภทกีฬา โดยลักษณะวิถีทัศนระดับล่างที่นำมาทดสอบได้แก่ ฮิสโตแกรมสี ออร์โตคอรัลโกลแกรม ลักษณะการเคลื่อนที่ของกล้อง และ ขอบสี

นอกจากนี้บทความทั้งหมดข้างต้นสังเกตได้ว่าผลความแม่นยำของการจำแนกไม่ได้เป็นปัจจัยหลักสำหรับระบบการจำแนกประเภทวิถีทัศนกีฬาเท่านั้น แต่ขนาดของปริภูมิเวกเตอร์ลักษณะเฉพาะยังเป็นอีกปัจจัยหนึ่งที่สำคัญซึ่งจะส่งผลถึงขนาดของข้อมูลในการจัดเก็บในการทำดัชนีและเวลาในการฝึกฝนสำหรับระบบการจำแนกประเภทวิถีทัศนกีฬา ดังนั้นในระบบการจำแนกประเภทวิถีทัศนกีฬาจำเป็นต้องมีการลดขนาดของปริภูมิลักษณะเฉพาะระดับล่าง โดยการเลือกปริภูมิที่มีนัยสำคัญเท่านั้น ซึ่งระบบการจำแนกวิถีทัศนกีฬาโดยทั่วไปจะอธิบายในหัวข้อที่ 1.3

1.3 ระบบการจำแนกประเภทวิถีทัศนกีฬาโดยใช้ลักษณะเฉพาะระดับล่างของวิถีทัศน

ระบบการจำแนกวิถีทัศนประเภทกีฬาโดยใช้ลักษณะเฉพาะระดับล่างของวิถีทัศนเป็นระบบการจำแนกประเภทกีฬาที่อาศัยลักษณะเฉพาะระดับล่างของวิถีทัศน เช่น สี รูปร่าง พื้นผิวภายในภาพ หรือ ลักษณะการเคลื่อนที่ของกล้อง และ/หรือ วัตถุ เป็นต้น เพื่อทำดัชนีหรือให้คำประกอบประเภทวิถีทัศนกีฬาที่ตรงกับความเข้าใจของมนุษย์ ในระบบการจำแนกโดยทั่วไปจะประกอบด้วยกระบวนการที่สำคัญ 2 กระบวนการ คือ กระบวนการเรียนรู้ (learning process) คือ การดึงลักษณะเฉพาะระดับล่างจากวิถีทัศน ซึ่งอยู่ในรูปแบบของค่าคงที่ที่เป็นตัวเลขที่เรียกว่าเวกเตอร์ลักษณะเฉพาะ เพื่อใช้เป็นข้อมูลตัวอย่างสำหรับการรู้จำแบบรูป (pattern recognition) และ กระบวนการจำแนก (Classification process) เมื่อผู้ใช้ต้องการจำแนกประเภทกีฬา ระบบจะทำการดึงเอาเวกเตอร์ลักษณะเฉพาะจากวิถีทัศนที่ต้องการจำแนกออกมา และ นำเวกเตอร์ลักษณะเฉพาะมาประมาณค่าความน่าจะเป็นในแต่ละกลุ่มวิถีทัศนกีฬา หรือนำมาเปรียบเทียบกับพารามิเตอร์ของกลุ่มวิถีทัศนกีฬา ซึ่งเทคนิคการหาค่าประมาณความน่าจะเป็นและการหาพารามิเตอร์ในการเปรียบเทียบได้มาจากกระบวนการเรียนรู้ เพื่อจำแนกประเภทวิถีทัศนกีฬาให้ตรงตามประเภทกีฬาได้มากที่สุด

1.4 เทคนิคที่นำเสนอ

ระบบการจำแนกวีดิทัศน์ประเภทกีฬาโดยใช้ลักษณะเฉพาะระดับล่างของวีดิทัศน์ จะมีประสิทธิภาพการจำแนกที่แม่นยำจะขึ้นกับการเลือกใช้เวกเตอร์ลักษณะเฉพาะที่ทำหน้าที่แทนประเภทกีฬาที่เหมาะสมและตัวจำแนกที่ได้จากการรู้จำแบบรูปที่เหมาะสมในกระบวนการเรียนรู้ ทั้งนี้เมื่อจำนวนประเภทของกีฬามีจำนวนกีฬามากขึ้น โอกาสของประเภทกีฬาที่มีเวกเตอร์ลักษณะเฉพาะคล้ายคลึงกันยิ่งมีมากขึ้นเช่นกัน ดังนั้นจึงมีแนวคิดในการวิเคราะห์แต่ละเวกเตอร์ลักษณะเฉพาะระดับล่างของวีดิทัศน์ เช่น ฮิสโตแกรมสี ออโตคอรีโโลแกรม ขอบสี และลักษณะเฉพาะของการเคลื่อนไหวของกล้อง อีกทั้งวิเคราะห์ตัวจำแนกที่เหมาะสมกับการจำแนกประเภทวีดิทัศน์กีฬาซึ่งเทคนิคที่นำเสนอประกอบด้วย โครงข่ายนิรอรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ เนื่องจากโครงข่ายนิรอรอลที่ไม่จำเป็นต้องมีขนาดใหญ่ก็สามารถนำมาใช้ในการจำแนกได้และให้ผลความแม่นยำที่ดีกว่าโครงข่ายนิรอรอลที่มีขนาดใหญ่ [48] นอกจากนี้ยังช่วยในการลดขนาดของปริภูมิเวกเตอร์ลักษณะเฉพาะ อีกทั้งโครงข่ายนิรอรอลเป็นตัวจำแนกที่ออกแบบได้ง่าย และให้ผลความแม่นยำการจำแนกที่ดีในหลายประยุกต์การใช้งาน ทั้งสะดวกต่อผู้ฝึกสอนเพียงกำหนดจำนวนอินพุตของโครงข่ายนิรอรอลเท่ากับขนาดปริภูมิของเวกเตอร์ลักษณะเฉพาะ และ จำนวนหน่วยนิรอรอลในชั้นเอาต์พุตเท่ากับประเภทของกีฬา สำหรับตัวจำแนกในปริภูมิสุญญ์ (null space based classifier) ตัวจำแนกในปริภูมิสุญญ์ได้จากการหาเมทริกซ์การฉายที่ทำให้การกระจายของข้อมูลระยะห่างระหว่างกลุ่มห่างกันมากที่สุด ในขณะที่ทำให้การกระจายของข้อมูลภายในกลุ่มเดียวกันมีระยะห่างใกล้กันมากที่สุด ซึ่งสามารถทำให้จำแนกข้อมูลได้ง่ายขึ้น

1.5 วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาเทคนิคการจำแนกประเภทวีดิทัศน์กีฬาโดยใช้ลักษณะระดับล่างของวีดิทัศน์ในระบบการค้นคืนวีดิทัศน์
2. เพื่อสร้างต้นแบบของระบบการจำแนกประเภทวีดิทัศน์กีฬา

1.6 ขอบเขตของการวิจัย

1. พัฒนาเทคนิคการสร้างตัวจำแนกประเภทวีดิทัศน์กีฬาได้แก่ ฟุตบอล บาสเก็ตบอล เทนนิส และ มวย โดยใช้ลักษณะระดับล่าง เช่น สี ขอบสี และ ลักษณะการเคลื่อนไหวที่ในระบบการค้นคืนวีดิทัศน์
2. พัฒนาโปรแกรมต้นแบบระบบการจำแนกประเภทวีดิทัศน์กีฬา

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. ความรู้เกี่ยวกับการจำแนกประเภทวีดิทัศน์กีฬาในระบบการค้นคืนวีดิทัศน์
2. โปรแกรมต้นแบบ เพื่อใช้จำแนกประเภทวีดิทัศน์กีฬา

1.8 วิธีดำเนินการวิจัย

1. สํารวจและทบทวนงานวิจัยที่เกี่ยวข้องกับระบบการจัดเก็บและค้นคืนภาพและวีดิทัศน์ที่ผ่านมาเพื่อเป็นแนวความคิดในงานวิจัย
2. ศึกษาโครงสร้างโดยรวมของระบบ เช่น การแยกส่วนวีดิทัศน์ทางเวลา การดึงเฟรมสำคัญ การดึงเวกเตอร์ลักษณะ การเปรียบเทียบความแตกต่างเวกเตอร์ลักษณะ การทำดัชนีวีดิทัศน์ การจำแนกวีดิทัศน์ และการวัดประสิทธิภาพ
3. นำเสนอและพัฒนาตัวจำแนกวีดิทัศน์ที่ใช้ลักษณะระดับล่าง เช่น สี ขอบ และการเคลื่อนที่ของกล้อง
4. วิเคราะห์ประสิทธิภาพของตัวจำแนกที่พัฒนาขึ้น
5. กำหนดแนวทางการวิจัยในอนาคต
6. สรุปงานวิจัยและเขียนวิทยานิพนธ์ฉบับสมบูรณ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ความรู้พื้นฐาน

เนื้อหาที่กล่าวในบทนี้เป็นทฤษฎีและความรู้พื้นฐานต่างๆ ที่เกี่ยวข้องกับวิทยานิพนธ์ ได้แก่ หัวข้อที่ 2.1 การแยกส่วนทางวีดิทัศน์ได้กล่าวถึงนิยามคำจำกัดความที่ใช้ในวิจัยเกี่ยวกับ การค้นคืนวีดิทัศน์ การให้คำประกอบวีดิทัศน์ การจำแนกวีดิทัศน์ เป็นต้น โครงสร้างของวีดิทัศน์โดยทั่วไป และ เทคนิคการแยกส่วนทางวีดิทัศน์ หัวข้อที่ 2.2 กล่าวถึงเทคนิคการหาเฟรมสำคัญเพื่อดึงลักษณะเฉพาะเชิงพื้นที่จากวีดิทัศน์ หัวข้อที่ 2.3 กล่าวถึงฉากวีดิทัศน์ที่พบที่เห็นได้จากการแพร่สัญญาณโทรทัศน์โดยทั่วไป หัวข้อที่ 2.4 เป็นการจำแนกข้อต่อเบื้องต้นเพื่อนำเอาข้อต่อที่มีเฉพาะการแข่งขันกีฬาเท่านั้นออกมา หัวข้อที่ 2.5 ลักษณะระดับล่างของวีดิทัศน์ ซึ่งประกอบด้วย ลักษณะเฉพาะฮิสโตแกรมสี ออโตคออริโลแกรม ลักษณะเฉพาะการเคลื่อนที่ของกล้อง และ ลักษณะเฉพาะขอบสีนำมาใช้ในวิทยานิพนธ์ และ หัวข้อที่ 2.6 เป็นหัวข้อสุดท้ายเป็นทฤษฎีเบื้องต้นที่ใช้ในการจำแนกประเภทวีดิทัศน์โดยใช้การรู้จำแบบรูป

2.1 การแยกส่วนวีดิทัศน์ทางเวลา (Temporal Video Segmentation)

ก่อนที่จะกล่าวในรายละเอียดของเทคนิควิธีการแยกส่วนวีดิทัศน์ทางเวลานั้น จะต้องนิยามคำจำกัดความที่ใช้ในงานวิจัยทางด้านนี้เสียก่อนเพื่อให้เกิดความเข้าใจที่ตรงกัน [17] คือ

ช็อตวีดิทัศน์ (video shot) คือ ลำดับของภาพที่มีความต่อเนื่องกันโดยที่ต้องบันทึกได้จากกล้องเพียงตัวเดียว ซึ่งถือว่าเป็นหน่วยย่อยที่สุดของข้อมูลวีดิทัศน์ ในที่นี้จะเรียกว่าช็อต

เฟรมสำคัญ (key frame) คือ ภาพที่ใช้แทนเนื้อหาโดยรวมของช็อต ในการหาเฟรมสำคัญนั้นอาจจะแทนช็อตด้วยภาพเดียว หรือ หลายภาพก็ได้ขึ้นอยู่กับเนื้อหาของช็อตมีความซับซ้อนเพียงใด

ฉากวีดิทัศน์ (video scene) คือ กลุ่มของช็อตที่ติดกันมีความสัมพันธ์กันทางด้านเวลาและเรื่องราว เพื่อใช้บรรยายเหตุการณ์ หรือ บอกเล่าเรื่อง ข้อแตกต่างระหว่างช็อตกับฉากวีดิทัศน์คือรอยต่อระหว่างช็อตจะเป็นรอยต่อทางกายภาพหรือก็คือรอยต่อที่ได้จากขั้นตอน

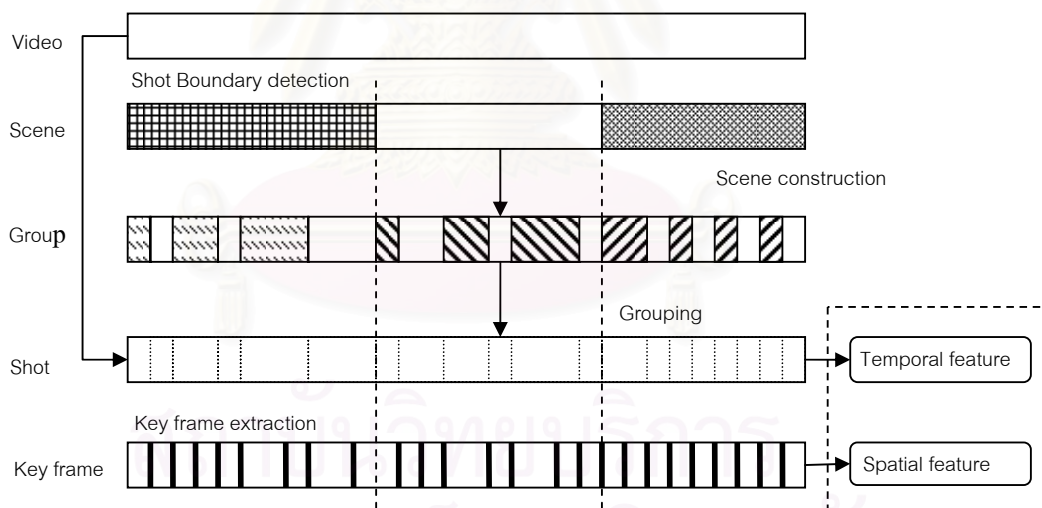
การผลิตวีดิทัศน์เรียกว่ารอยต่อทางกายภาพ (physical boundary) ในขณะที่รอยต่อที่เกิดจากฉากวีดิทัศน์เป็นรอยต่อทางด้านเรื่องราวของวีดิทัศน์

กลุ่มวีดิทัศน์ (video group) เป็นกลุ่มของช็อตที่มีความคล้ายคลึงกัน หรือ มีความเหมือนกันเพื่อใช้เชื่อมต่อระหว่างช็อตกับฉากวีดิทัศน์

คำจำกัดความต่างๆ ที่ใช้ในงานวิจัยทางด้านนี้สามารถเขียนเป็นโครงสร้างได้ 5 ระดับ คือ ระดับวีดิทัศน์ ระดับฉาก ระดับกลุ่ม ระดับช็อต และ ระดับเฟรมสำคัญ ดังรูปที่ 2.1

2.1.1 ประเภทของการเปลี่ยนแปลงฉาก

จากรูปที่ 2.1 ข้อมูลวีดิทัศน์ซึ่งเป็นลำดับภาพที่มีความยาวประกอบด้วยหลายช็อตนำมาต่อกันจะทำให้เกิดรอยต่อระหว่างช็อต (shot boundary) โดยเกิดจากการผลิตข้อมูลวีดิทัศน์ซึ่งเป็นรอยต่อทางกายภาพไม่ใช่รอยต่อของฉาก (scene boundary) ในการจัดประเภทรอยต่อทางกายภาพสามารถจัดได้เป็น 2 แบบ [18] คือ การเปลี่ยนแปลงฉากแบบทันที (abrupt transition) และ การเปลี่ยนแปลงฉากแบบค่อยเป็นค่อยไป (gradual transition)



รูปที่ 2.1 โครงสร้างระดับวีดิทัศน์

การเปลี่ยนแปลงฉากแบบทันที เป็นรอยต่อที่เกิดขึ้นโดยนำสองช็อตมาวางติดกันทำให้ช็อตทั้งสองไม่มีความเกี่ยวข้องซึ่งกัน ดังรูปที่ 2.2



รูปที่ 2.2 รอยต่อของข้อตแบบการเปลี่ยนแปลงทันที

การเปลี่ยนแปลงจากแบบค่อยเป็นค่อยไป เป็นรอยต่อที่เกิดจากเพิ่มเทคนิคพิเศษบางอย่างเพื่อนำสองข้อตมาวางติดกันซึ่งระหว่างสองข้อตที่มาวางติดกันจะมีความสัมพันธ์กัน ซึ่งแบ่งได้อีก 3 รูปแบบคือ

- แบบเฟดเข้า หรือ เฟดออก (Fade-in/Fade-out) เป็นรอยต่อของข้อตที่มีการเปลี่ยนค่าความส่องสว่างของแต่ละเฟรมเพื่อสร้างรอยต่อซึ่งแสดงให้เห็นดังรูปที่ 2.3



รูปที่ 2.3 การเปลี่ยนแปลงจากแบบเฟดออก

- แบบเฟดซ้อน (Dissolve) เป็นรอยต่อที่เกิดจากสองข้อตที่ติดกันโดยที่ข้อตหนึ่งมีการเปลี่ยนแปลงจากแบบเฟดออกขณะที่อีกข้อตมีการเปลี่ยนแปลงจากแบบเฟดเข้าดังรูปที่ 2.3



รูปที่ 2.4 การเปลี่ยนแปลงจากแบบเฟดซ้อน

- การปาด (Wipe) เป็นรอยต่อของข้อตโดยที่จะเห็นขอบเฟรมของข้อตเก่าในรูปแบบใดรูปแบบหนึ่งดังรูปที่ 2.5



รูปที่ 2.5 การเปลี่ยนแปลงจากแบบปาด

2.1.2 เทคนิคการแยกส่วนวีดิทัศน์ทางเวลา

การแยกส่วนวีดิทัศน์ทางเวลา เป็นการแยกส่วนของวีดิทัศน์ที่ประกอบด้วยหลายช็อตที่นำมาต่อกันเพื่อหาตำแหน่งรอยต่อระหว่างช็อตภายในวีดิทัศน์ ซึ่งที่เทคนิคการแยกส่วนทางวีดิทัศน์ทางเวลาดังต่อไปนี้

2.1.2.1 เทคนิคผลต่างจุดภาพ [18] (Pixel Difference Technique)

เทคนิคนี้เป็นการหาค่าผลต่างระหว่างสองเฟรมที่ติดกันจากจุดภาพ โดยค่าผลต่างระหว่างเฟรมได้จากผลรวมความแตกต่างของค่าความเข้มแสง หรือ ค่าของสี ในจุดภาพระหว่างสองเฟรมซึ่งเทคนิคการหาค่าผลต่างระหว่างเฟรมเทคนิคนี้สามารถหาการเปลี่ยนแปลงฉากแบบทันทีได้โดยการใช้ขีดเริ่มเปลี่ยนเป็นตัวตรวจหาการเปลี่ยนแปลง

2.1.2.2 เทคนิคผลต่างฮิสโตแกรม (Histogram Difference Technique)

เนื่องจากเทคนิคจุดภาพเมื่อมีการเคลื่อนที่ของวัตถุหรือการเคลื่อนที่ของกล้องทำให้ผลของค่าความแตกต่างมีมากขึ้น จึงทำให้การแยกส่วนวีดิทัศน์มีความผิดพลาด จึงมีการนำเสนอ การใช้ฮิสโตแกรมซึ่งเป็นการหาจำนวนของระดับความเข้มของสีว่ามีค่าความถี่เป็นเท่าใดเพื่อที่จะลดผลกระทบที่เกิดจากการเคลื่อนที่ของกล้องและการเคลื่อนที่ของวัตถุจึงมีการเสนอวิธีการโดยใช้เทคนิคของการเปรียบเทียบฮิสโตแกรมระหว่างสองเฟรมที่ติดกัน โดยที่แนวคิดนี้เกิดจากพื้นหลังและวัตถุเดียวกันซึ่งอยู่ในเฟรมที่ติดกันจะมีการเคลื่อนที่เพียงเล็กน้อยเท่านั้น ดังนั้นฮิสโตแกรมระหว่างสองเฟรมที่ติดกันจะมีรูปร่างคล้ายกัน และ ให้ค่าผลต่างเพียงเล็กน้อย นอกจากนี้คุณสมบัติที่สำคัญอีกอย่างหนึ่งคือรูปร่างของฮิสโตแกรมจะไม่เปลี่ยนแปลงเมื่อมีการหมุนของภาพเกิดขึ้น หรือ เมื่อภาพมีการหมุน ซึ่งเทคนิคการหาผลต่างโดยใช้ฮิสโตแกรม ได้แก่

- การเปรียบเทียบฮิสโตแกรมรวม (Global Histogram Comparison)

วิธีการที่ง่ายที่สุดในการเปรียบเทียบโดยใช้ฮิสโตแกรมดังสมการที่ (2.1)

$$D(i, i+1) = \sum_{j=1}^n |H_i(j) - H_{i+1}(j)| \quad (2.1)$$

โดยที่ $H_i(j)$ เป็นค่าฮิสโตแกรมสำหรับระดับเทา j ในเฟรมที่ i ซึ่ง j คือค่าสีเทาในขณะที่ n เป็นจำนวนของระดับสีเทาโดยทั่วไปแล้วเป็น 256 ระดับ (0-255) เพื่อปรับปรุงให้ค่าผลต่างมีค่าสูงขึ้นเมื่อมีการเปลี่ยนแปลงฉากแบบทันที ดังนั้นจึงได้มีการนำเสนอโดยการยกกำลังสองเพื่อเพิ่มค่าผลต่างขึ้นเรียกว่า χ^2 -test (Chi Square Test) [2] ดังสมการที่ (2.2)

$$D(i, i+1) = \sum_{j=0}^n \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)} \quad (2.2)$$

โดยที่ $H_i(j)$ เป็นค่าฮิสโตแกรมของรหัสสี j ณ เฟรมที่ i โดยที่รหัสสีมีจำนวนเป็น $n+1$ รหัสสี เทคนิคฮิสโตแกรมแบบนี้ๆ เพื่อใช้ในการหาค่าผลต่างระหว่างสองเฟรมมีดังต่อไปนี้

- ฮิสโตแกรมอินเตอร์เซกชัน (Histogram Intersection) [19]

$$D(i, i+1) = 1 - \frac{\sum_{k=0}^n \min(H_i(k), H_{i+1}(k))}{\sum_{k=0}^n \max(H_i(k), H_{i+1}(k))} \quad (2.3)$$

เมื่อสองเฟรมที่ติดกันมีฮิสโตแกรมที่เหมือนกันและนำมาอินเตอร์เซกชันกันผลที่ได้จะเป็น 1 ในขณะที่เมื่อแตกต่างกันแล้วผลที่ได้จะเป็น 0

- ค่าต่างของถั่งสีแบบถ่วงน้ำหนัก (Weighted bin difference) [19]

$$D(i, i+1) = \sum_{j=1}^n \sum_{k \in N(k)} W(k) \cdot (H_i(j) - H_{i+1}(k)) \quad (2.4)$$

โดยที่ $N(k)$ เป็นบริเวณใกล้เคียง (Neighborhood) ของ รหัสสี j และ $W(k)$ เป็นค่าน้ำหนักที่ถ่วงให้กับแต่ละบริเวณใกล้เคียง

2.1.2.3 เทคนิคผลต่างแบบบล็อก (Block difference techniques) [20]

จากเทคนิคในหัวข้อ 2.1.2.1 และ หัวข้อ 2.1.2.2 เป็นเทคนิคหาค่าผลต่างของทั้งภาพสำหรับเทคนิคหาค่าผลต่างแบบบล็อกนี้เป็นเทคนิคเพื่อลดผลอันเนื่องมาจากสัญญาณรบกวน และ ผลเนื่องมาจากการแวบสว่าง (Flash) โดยการกระจายภาพออกเป็นส่วนย่อยเรียกว่า บล็อกซึ่งภาพในเฟรมที่ i จะถูกแบ่งออกเป็น r บล็อกโดยที่บล็อกแต่ละบล็อกมีตำแหน่งเดียวกันระหว่างเฟรมที่ i และ $i+1$ และค่าผลต่างคำนวณได้จากสมการที่ (2.5)

$$D(i, i+1) = \sum_{k=1}^r C_k \times DB(i, i+1, k) \quad (2.5)$$

เมื่อ C_k เป็นค่าถ่วงน้ำหนัก และ $DB(i, i+1, k)$ เป็นค่าผลต่างของการจับคู่ระหว่างบล็อกที่ตรงกันระหว่างเฟรมที่ i และ $i+1$

- ค่าผลต่างแบบบล็อกโดยเทคนิคจุดภาพ

Kasturi และ Jain ได้นำเสนอการเปรียบเทียบค่า Likelihood ratio โดยใช้คุณสมบัติทางสถิติของค่าเข้มแสงในแต่ละบล็อกที่ตรงกันดังสมการที่ (2.6)

$$\text{Likelihood ratio} = \frac{\left(\frac{\mu_{i,k} + \mu_{i+1,k}}{2} + \frac{\sigma_{i,k}^2 - \sigma_{i+1,k}^2}{2} \right)^2}{\sigma_{i,k}^2 \times \sigma_{i+1,k}^2} \quad (2.6)$$

โดยที่ $\mu_{i,k}$ และ $\sigma_{i,k}^2$ เป็นค่าเฉลี่ยและค่าความแปรปรวนของแต่ละบล็อกตามลำดับ เมื่อค่าที่ได้จากสมการที่ (2.6) มีค่ามากกว่าขีดเริ่มเปลี่ยนค่าหนึ่ง แล้วบล็อกจะถูกกำหนดให้มีความเป็นหนึ่ง ซึ่งหมายความว่าบล็อก ณ ตำแหน่งนั้นมีความแตกต่างกัน แต่ถ้าค่าที่ได้น้อยกว่าแล้วจะกำหนดให้มีความเป็นศูนย์ซึ่งหมายความว่าบล็อก ณ ตำแหน่งนั้นไม่แตกต่างกันหรือเหมือนกัน จากสมการที่ (2.5) ถ้ากำหนดให้ C_k มีค่าเป็นหนึ่งแล้วถือว่าการนับจำนวนบล็อกที่เปลี่ยนแปลงระหว่างเฟรมซึ่งเป็นค่าผลต่างระหว่างเฟรม

- ค่าผลต่างแบบบล็อกโดยเทคนิคฮิสโตแกรม

Swanberg และ คณะ นำข้อดีจากการแบ่งบล็อกมาใช้กับฮิสโตแกรมดังสมการที่ (2.7) และ กำหนดให้ $C_k = 1/r$ สำหรับทุกบล็อก

$$DB(i, i+1, k) = \sum_{c \in \{R, G, B\}} \sum_{j=0}^{256} \frac{(H_c(i, j) - H_c(i+1, j))^2}{H_c(i, j) + H_c(i+1, j)} \quad (2.7)$$

2.1.3 การเลือกขีดเริ่มเปลี่ยนเพื่อตรวจหารอยต่อของช็อต (Shot boundary detection)

เมื่อได้ค่าผลต่างระหว่างเฟรมแล้วเพื่อที่จะตรวจหาการเปลี่ยนแปลงของฉากในข้อมูลวีดิทัศน์จึงจำเป็นต้องกำหนดขีดเริ่มเปลี่ยนโดยเมื่อค่าผลต่างระหว่างเฟรมมีค่ามากกว่าขีดเริ่มเปลี่ยนแล้วถือวาระหว่างเฟรมนั้นเกิดการเปลี่ยนแปลงฉากขึ้น โดยในที่นี้เป็นการตรวจหาสำหรับการเปลี่ยนแปลงฉากแบบทันที สำหรับการเลือกขีดเริ่มเปลี่ยนนี้เป็นไปได้ 2 แนวทางคือ เลือกขีดเริ่มเปลี่ยนแบบโดยรวม และ เลือกขีดเริ่มเปลี่ยนแบบปรับตัวได้

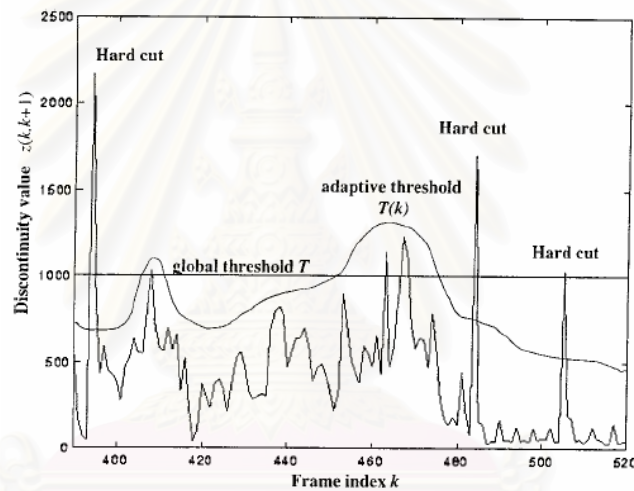
2.1.3.1 ขีดเริ่มเปลี่ยนแบบโดยรวม (Global Threshold) [21]

เป็นการวัดค่าการกระจายทางสถิติของค่าผลต่างของจุดภาพภายในเฟรมวีดิทัศน์ และ จำลองค่าการกระจายตัวทางสถิติของค่าผลต่างให้อยู่ในรูปแบบฟังก์ชันเกาส์เซียน ซึ่งมี

พารามิเตอร์ μ คือ ค่าเฉลี่ยของผลต่าง และ พารามิเตอร์ σ เป็นค่าเบี่ยงเบนมาตรฐาน ของผลต่างโดยที่ขีดเริ่มเปลี่ยนหาได้จากสมการที่ (2.8)

$$T = \mu + r\sigma \quad (2.8)$$

เมื่อ T เป็นขีดเริ่มเปลี่ยน และ r เป็นพารามิเตอร์ที่กำหนดขึ้นเพื่อบอกความสัมพันธ์ของโอกาสความทนทานต่อการตรวจหารอยต่อของขีดที่ผิดพลาด ตัวอย่างเช่น เมื่อ r เท่ากับ 3 โอกาสที่มีการตรวจหาผิดพลาดเป็น 0.1% เป็นต้น ข้อเสียของการเลือกขีดเริ่มเปลี่ยนแบบรวม คือ เมื่อกำหนดให้ขีดเริ่มเปลี่ยนให้มีค่าน้อยเกินไปจะทำให้โอกาสที่ตรวจหาผิดพลาด (false detection) มีมากขึ้น ในขณะที่ถ้าขีดเริ่มเปลี่ยนมีค่าสูงเกินไปเพื่อที่จะหลีกเลี่ยงการตรวจหาที่ผิดพลาดอาจจะทำให้ตรวจหาไม่พบ (miss detection) ได้ดังรูปที่ 2.6



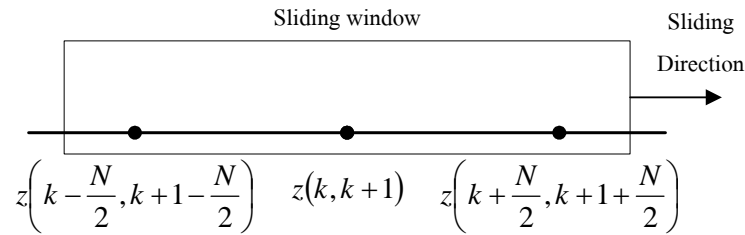
รูปที่ 2.6 การหาขีดเริ่มเปลี่ยนใจแบบรวม (T) และ แบบปรับตัวได้ ($T(k)$) [5]

2.1.3.2 ขีดเริ่มเปลี่ยนแบบปรับตัวได้ (Adaptive threshold) [5]

เป็นการปรับปรุงขีดเริ่มเปลี่ยนให้ดีขึ้น ซึ่งจากรูปที่ 2.6 แสดงให้เห็นว่าการเลือกขีดเริ่มเปลี่ยนแบบปรับตัวได้ให้ได้ผลดีกว่าการเลือกขีดเริ่มเปลี่ยนโดยรวม วิธีการหาขีดเริ่มเปลี่ยนนั้นหาได้จากค่าผลต่างที่อยู่ใกล้เคียงในรูปแบบของหน้าต่างเลื่อน (sliding window) แสดงดังรูปที่ 2.7 การตรวจหารอยต่อของข้อต่อมีเงื่อนไขที่ว่า

$$T(k) = \frac{\mu}{k - \frac{N}{2} < k < k + \frac{N}{2}}(k) + r \cdot \frac{\sigma}{k - \frac{N}{2} < k < k + \frac{N}{2}}(k) \quad (2.9)$$

จากสมการที่ (2.9) โดยที่ $T(k)$ เป็นขีดเริ่มเปลี่ยนใจในแต่ละเฟรมที่ k ในหน้าต่างขนาด N เฟรม และ r เป็นพารามิเตอร์เดียวกับสมการที่ (2.8)



รูปที่ 2.7 วิธีการหาขีดเริ่มเปลี่ยนใจโดยใช้หน้าต่างเคลื่อน [5]

2.2 การดึงเฟรมสำคัญ (Key Frame Extraction)

เพื่อที่จะสร้างฐานข้อมูลลักษณะ ถ้านำเฟรมในแต่ละข้อต่อวิดีโอที่นำมาดึงลักษณะของแต่ละเฟรมออกมา ผลที่ได้คือจำนวนเฟรมสำคัญในฐานข้อมูลมีจำนวนมากเกินไป ทั้งนี้เนื่องจากเฟรมที่ติดกันในบางข้อต่อวิดีโอจะมีลักษณะที่คล้ายกัน ดังนั้นระบบการค้นคืนวิดีโอจึงจำเป็นต้องมีขั้นตอนที่เรียกว่าการคัดย่อวิดีโอ (video abstraction) ในบางบทความเรียกว่าบทสรุปวิดีโอ (video summarization)

การคัดย่อวิดีโอ [22] คือ การทำให้วิดีโอมีจำนวนเฟรมที่น้อยลงหรือกระชับเพื่อให้แทนข้อมูลวิดีโอ โดยนำมาใช้ในการประยุกต์การใช้งานวิดีโอหลากหลาย ตัวอย่างเช่น การเข้าถึงแบบสุ่ม (random access) เพื่อเข้าถึงจุดของวิดีโอที่ต้องการอย่างรวดเร็ว และ ใช้ในการสร้างระบบค้นคืนวิดีโอ ซึ่งการคัดย่อวิดีโอสามารถแบ่งได้เป็น 2 รูปแบบคือ

ลำดับตัวอย่าง (preview sequence) [22] ถือว่าเป็นการแยกหาวิดีโอที่สำคัญ (Key Video Segmentation) คือ การลดลำดับภาพวิดีโอที่เป็นลำดับที่ยาวให้เป็นลำดับภาพที่สั้นลงซึ่งลำดับภาพที่สั้นลงนี้ถือว่าเป็นส่วนของวิดีโอที่มีความสำคัญมากที่สุด

เฟรมสำคัญ (key frame) เป็นเซตของเฟรมในวิดีโอที่เลือกมาเพื่อเป็นตัวแทนของวิดีโอที่มีความสำคัญมากที่สุด

ในระบบการค้นคืนวิดีโอ การดึงเฟรมสำคัญมีความจำเป็นอย่างมาก เนื่องจากเฟรมสำคัญเป็นตัวแทนของข้อต่อวิดีโอที่นำไปใช้ในเพื่อทำดัชนีวิดีโอ ถ้าเฟรมที่สำคัญที่เลือกมาไม่ตรงกับสาระสำคัญที่สุดในข้อต่อวิดีโอ จะทำให้ผู้ใช้งานไม่สามารถค้นคืนวิดีโอได้ตามความต้องการ และ นอกจากนี้เพื่อให้ง่ายต่อการสร้างระบบการค้นคืนวิดีโอจึงได้ประยุกต์เทคนิคการหาเวกเตอร์ลักษณะเฉพาะระดับล่างของแต่ละภาพจากระบบการค้นคืนภาพ (Image Retrieval) เพื่อนำมาบ่งบอกเฟรมสำคัญ ดังนั้นเฟรมสำคัญถือว่าเป็นภาพในฐานข้อมูลซึ่งถ้าภาพ

สอบถามคล้ายคลึงกับเฟรมสำคัญใด ก็ถือว่า ข้อตววิทัศน์ที่มีเฟรมสำคัญนั้นเป็นผลที่ได้จากสอบถาม ซึ่งเทคนิคประกอบด้วยกัน 4 เทคนิคหลักดังต่อไปนี้

2.2.1 แนวความคิดแบบรอยต่อของช็อต (Shot Boundary based approach) [17]

หลังจากวิทัศน์ผ่านขั้นตอนของการแยกวิทัศน์ทางเวลาแล้วจะได้หน่วยย่อยที่เรียกว่าช็อตวิทัศน์ วิธีการอย่างง่ายในการดึงเฟรมสำคัญออกมาคือนำเฟรมแรกของแต่ละช็อตวิทัศน์มาเป็นเฟรมสำคัญ ถึงแม้ว่าวิธีการนี้เป็นวิธีการที่ง่ายแต่เฟรมสำคัญในแต่ละช็อตวิทัศน์ถูกจำกัดที่เพียงเฟรมเดียว จึงไม่สามารถครอบคลุมเนื้อหาของช็อตวิทัศน์ที่ลักษณะซับซ้อนได้ เช่น ในกรณีที่ช็อตวิทัศน์มีการเคลื่อนที่ของกล้องเกิดขึ้นจะทำให้ไม่สามารถใช้เฟรมแรกเพียงเฟรมเดียวเป็นเฟรมสำคัญที่แทนทั้งช็อตวิทัศน์ได้

2.2.2 แนวความคิดแบบเนื้อหาด้วยการมองเห็น (Visual Content Based approach) [24]

Zhang และ คณะ ได้นำเสนอหลายกฎเกณฑ์การมองเห็นเพื่อที่จะดึงเฟรมสำคัญคือ แทนที่จะใช้เพียงเฟรมแรกเฟรมเดียวของแต่ละช็อตวิทัศน์ ก็เลือกหลายเฟรมสำคัญจากช็อตวิทัศน์ เรียกว่า เซตของเฟรมสำคัญ (Set of Key frame) ซึ่งเป็นไปตามกฎเกณฑ์ดังต่อไปนี้

- กฎเกณฑ์พื้นฐานของช็อต (Shot based criteria) คือ การดึงเฟรมสำคัญในระดับช็อต โดยในแต่ละเฟรมแรกถือว่าเป็นเฟรมสำคัญ ส่วนเฟรมสำคัญที่เหลือขึ้นอยู่กับการวิเคราะห์ในกฎเกณฑ์ต่อมา

- กฎเกณฑ์พื้นฐานของลักษณะของสี (Color feature based criteria) คือ หลังจากได้เฟรมแรกของช็อตวิทัศน์ที่เป็นเฟรมสำคัญแล้ว จะเปรียบเทียบความเหมือนกันระหว่างเฟรมแรกกับเฟรมที่เหลือ ซึ่งมีขั้นตอนดังต่อไปนี้

$$\begin{aligned} \text{ขั้นตอนแรก} \quad & F_{last} = 1 \\ \text{ขั้นตอนที่ 2} \quad & \forall k \in [2, S] \quad \text{if } z(F_{last}, k) > T \Rightarrow F_{last} = k \end{aligned} \quad (2.10)$$

โดยที่ F_{last} เป็นเฟรมสำคัญสุดท้ายที่ดึงมาจากช็อตวิทัศน์เดียวกัน ซึ่งในขั้นตอนแรกเฟรมสุดท้ายคือเฟรมแรกของช็อตวิทัศน์ และ $z(F_{last}, k)$ คือค่าที่ได้จากการคำนวณหาค่าเปรียบเทียบความแตกต่างระหว่าง k ซึ่งเป็นเฟรมปัจจุบันในลำดับภาพช็อตวิทัศน์กับเฟรมสำคัญสุดท้าย F_{last} โดยใช้วิธีฮิสโตแกรมสี หรือ โมเมนต์สี (color moment) ซึ่งจากสมการที่ (2.10) อธิบายได้ว่าทุกลำดับภาพที่ 2 ถึง S ซึ่งเป็นจำนวนลำดับภาพของช็อตวิทัศน์ ถ้าลำดับภาพที่ k กับเฟรมสำคัญสุดท้ายมีความแตกต่างกันมากกว่าขีดเริ่มเปลี่ยน T แล้ว ลำดับภาพที่

k เป็นเฟรมสำคัญใหม่และ แทน F_{last} ด้วยลำดับภาพที่ k ซึ่งในขั้นตอนที่ 2 จะเปรียบเทียบค่าไปจนกระทั่งเฟรมสุดท้ายของข้อ

- กฎเกณฑ์พื้นฐานของการเคลื่อนที่ เป็นการปรับปรุงโดยใช้ข้อมูลของการเคลื่อนที่โดยรวม (global motion) ที่ได้จากการดำเนินการของกล้อง (camera operation) และการเคลื่อนที่ของวัตถุที่ขนาดใหญ่ในข้อตวิทัศน์ เนื่องจากการดำเนินการของกล้อง ใช้เพื่อตรวจหาเหตุการณ์ หรือ ต้องการเน้นวัตถุอย่างชัดเจน ดังนั้น กฎเกณฑ์พื้นฐานของลักษณะของสีไม่สามารถตรวจหาค่าการเคลื่อนที่นี้ได้ จึงจำเป็นต้องมีการตรวจหาการเคลื่อนที่รวมในแนวหลัก 2 แนวคือ การถ่าย (panning) และ การซูม (zooming) โดยข้อตวิทัศน์ที่มีการซูมจะต้องมีเฟรมสำคัญอย่างต่ำสองเฟรมคือ เฟรมแรกกับเฟรมสุดท้ายของการซูม ส่วนในกรณีการถ่ายนั้นจำนวนของเฟรมสำคัญจะขึ้นอยู่กับอัตราการเปลี่ยนแปลงในการมองเห็นเนื้อหาภาพ นั่นคือเฟรมสำคัญที่ได้จากการถ่ายจะต้องมีการซ้อนทับกันน้อยที่สุด หรือ วัตถุในแต่เฟรมสำคัญเป็นกิจกรรมของวัตถุที่แตกต่างกัน โดยทั่วไปแล้วเฟรมสำคัญควรมีเนื้อหาภายในภาพน้อยกว่า 30%

2.2.3 แนวความคิดแบบการวิเคราะห์การเคลื่อนที่ (Motion analysis based approach) [25]

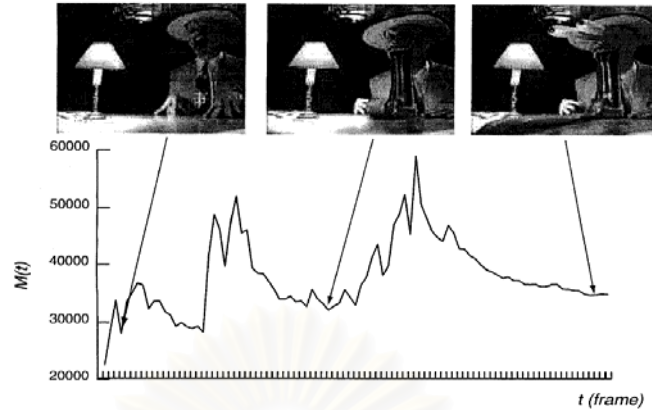
Wolf ได้นำเสนอการดึงเฟรมสำคัญโดยพิจารณาจากการเคลื่อนที่ กล่าวคือใช้การวิเคราะห์การไหลเชิงแสง (Optical Flow Analysis) เพื่อที่จะวัดค่าการเคลื่อนที่ที่เกิดขึ้นในข้อตวิทัศน์ และ เลือกเฟรมสำคัญจากช่วงที่มีค่าของการเคลื่อนที่น้อยที่สุดซึ่งช่วงนี้เรียกว่าค่าที่น้อยที่สุดของช่วง (Local minima) โดยระเบียบวิธีการของ Wolf มีขั้นตอนดังต่อไปนี้

คำนวณหาค่าการไหลเชิงแสงในแต่ละเฟรมแล้วคำนวณหาเมตริกซ์อย่างง่ายของการเคลื่อนที่ดังสมการที่ (2.11)

$$M(t) = \sum_i \sum_j |o_x(i, j, t)| + |o_y(i, j, t)| \quad (2.11)$$

วิธีการหาค่าการไหลเชิงแสง Wolf ใช้ระเบียบวิธีการของ Horn และ Schunck ซึ่งจะได้ $o_x(i, j, t)$ เป็นองค์ประกอบในแกน x ที่ตำแหน่งจุดภาพ i และ j ณ เฟรมที่ t และ $o_y(i, j, t)$ เป็นองค์ประกอบในแกน y ที่ตำแหน่งจุดภาพ i และ j ณ เฟรมที่ t

วิเคราะห์เมตริกซ์ที่ได้จากการคำนวณเป็นฟังก์ชันของเวลา และ เลือกเฟรมสำคัญจากค่าการเคลื่อนที่ที่ต่ำที่สุด ระเบียบวิธีการนี้จะสแกน $M(t)$ ไปตามแกน t ซึ่งอธิบายได้ดังรูปที่ 2.8



รูปที่ 2.8 การดึงเฟรมสำคัญโดยระเบียบวิธีการของ Wolf [25]

จากรูปที่ 2.8 มีการกำหนดค่า 2 ค่าคือ m_1 และ m_2 เป็นช่วงของค่าที่มากที่สุดค่า (Local maxima) โดยที่ค่า m_2 จะน้อยกว่าค่า m_1 เป็น $N\%$ ในฟังก์ชัน $M(t)$ โดยที่ค่าต่ำสุดที่อยู่ระหว่าง m_1 และ m_2 จะถูกเลือกเป็นเฟรมสำคัญ ในการหาเฟรมสำคัญในช่วงถัดไป m_2 จากการหาเฟรมสำคัญครั้งที่แล้วจะแทนเป็น m_1 และค่า m_2 จะน้อยกว่าค่า m_1 เป็น $N\%$ ไปเรื่อยๆ ผลที่ได้แสดงให้เห็นดังรูปที่ 2.8 แสดงการดึงเฟรมสำคัญ 3 เฟรมออกมาจากลำดับวิดีโอ

2.2.4 แนวความคิดแบบการจัดกลุ่ม (Clustering based approach) [24]

Zhuang และ คณะ ได้เสนอวิธีการจัดกลุ่ม (Clustering) ที่ใช้ในงานการรู้จำรูปแบบ (Pattern Recognition) เพื่อที่จัดกลุ่มเฟรมในวิดีโอเพื่อหาเฟรมสำคัญ โดยให้ชุดวิดีโอ $s = \{f_1, f_2, \dots, f_N\}$ ได้มาจากขั้นตอนการแยกวิดีโอทางเวลา โดยที่ชุดวิดีโอจำนวน N เฟรม จัดให้อยู่ใน M กลุ่ม เรียกว่า $\sigma_1, \sigma_2, \dots, \sigma_M$ ในวิธีการนี้วัดความเหมือนกันระหว่างสองเฟรมโดยใช้ 16x8 2D HS color histogram intersection ในแบบจำลองสี HSV ดังสมการที่ (2.12)

$$sim(i, j) = \sum_{h=1}^{16} \sum_{s=1}^8 \min(H_i(h, s), H_j(h, s)) \quad (2.12)$$

โดยที่ สมการที่ (2.12) เป็นการค่าหาความเหมือนกันระหว่างสองเฟรม i และ j เมื่อ $H(h, s)$ เป็นฮิสโตแกรมสองมิติของแกน h (Hue) และ แกน s (Saturation) ในแบบจำลองสี HSV โดยในแกน h ถูกลดระดับของสีเหลือ 16 ระดับ ในทำนองเดียวในแกน s ถูกลดระดับของสีเหลือ 8 ระดับ โดยระเบียบวิธีการจัดกลุ่มจะมีขีดเริ่มเปลี่ยน δ ค่าหนึ่งเพื่อที่ควบคุมความหนาแน่นของการจัดกลุ่ม ถ้าค่า δ มีค่ามากจะทำให้จำนวนกลุ่มของการจัดกลุ่มมีมาก ซึ่งอธิบายได้ว่าถ้าวัดค่าความเหมือนกันระหว่างสองเฟรมที่ติดกันตามสมการที่ (2.12) แล้วมีค่าน้อยกว่าค่า δ เฟรมนั้นถือว่าเป็นเฟรมไม่คล้ายคลึงกันซึ่งอาจจะเป็นเฟรมสำคัญ

หลังจากจัดกลุ่มให้แต่ละเฟรมแล้วขั้นต่อไปคือการเลือกหาเฟรมสำคัญ โดยที่กลุ่มที่มีขนาดใหญ่ที่สุดเพียงพอที่จะเป็นกลุ่มสำคัญ (key cluster) มากกว่าค่าขนาดเฉลี่ยกลุ่ม N/M จะมีเฟรมสำคัญเพียงเฟรมเดียวที่ได้จากกลุ่มที่ใหญ่ที่สุด ถ้ามีขนาดน้อยกว่าค่าเฉลี่ยกลุ่มเฟรมสำคัญหาได้จากเฟรมที่มีค่าใกล้เคียงกับค่าศูนย์ถ่วงของแต่ละกลุ่ม

2.3 ฉากวีดิทัศน์กีฬา

การบันทึกวีดิทัศน์กีฬาประเภทต่างๆ ที่ได้จากการแพร่สัญญาณโทรทัศน์ วีดิทัศน์กีฬาที่ได้มีหลากหลายรูปแบบอาทิเช่นประเภทกีฬาในร่มหรือเป็นกลางแจ้งก็ได้ จำนวนนักกีฬาประเภทบุคคลเดียว หรือ ประเภททีม เป็นต้น ในวีดิทัศน์หนึ่งๆ จะมีความหลากหลายของเหตุการณ์ที่แตกต่างกันไป ทำให้ไม่สามารถกำหนดเงื่อนไขให้ผู้บรรยายหรือบทการสนทนาได้ ทั้งนี้ขึ้นอยู่กับปัจจัยภายนอกต่างๆ อีกด้วย เช่น เทคนิคการบันทึกหรือตัดต่อลำดับวีดิทัศน์ที่แตกต่างกันในแต่ละสถานีโทรทัศน์ และ ข้อตกลงระหว่างสถานีแพร่สัญญาณโทรทัศน์ด้วย แต่วีดิทัศน์กีฬาโดยทั่วไปจะมีฉากเหตุการณ์ที่เหมือนกันดังต่อไปนี้ [2]

- **ฉากโรงฉายหรือฉากสัมภาษณ์ (studio/interview)** เป็นลำดับข้อตวีดิทัศน์ที่แสดงการวิพากษ์วิจารณ์ของนักข่าวหรือผู้เชี่ยวชาญในกีฬาประเภทนั้น ส่วนใหญ่มักบันทึกวีดิทัศน์จากภายในโรงถ่าย
- **ฉากวัตถุกราฟิก (graphic object)** เป็นลำดับข้อตวีดิทัศน์ที่แสดงถึงตัวอักษรหรือรูปภาพเพื่อบ่งบอกรายละเอียดของการแข่งขัน เช่น ชื่อนักกีฬา ตารางคะแนน หรือสัญลักษณ์การแข่งขัน เป็นต้น
- **ฉากแข่งขัน (playing)** เป็นลำดับข้อตวีดิทัศน์ที่แสดงการแข่งขันกีฬา ซึ่งแตกต่างกันไปตามประเภทของกีฬา เราสามารถแบ่งข้อตในฉากแข่งขันได้เป็น ข้อตสนามแข่งขัน (Playing field shot) ซึ่งแสดงถึงการแข่งขันของนักกีฬานสนามการแข่งขัน โดยสนามแข่งขันจะเป็นสนามที่มีสีเป็นเนื้อเดียวแล้วตัดด้วยเส้นแบ่งบริเวณสนามแข่งขัน ข้อตนักกีฬา เป็น ข้อตที่ผู้ถ่ายต้องการเน้นไปยังตัวนักกีฬาเป็นฉากหน้า และ ฉากหลังของภาพจะมีลักษณะเป็นสีเนื้อเดียวหรือพัว (blur) เนื่องจากการเคลื่อนไหวของกล้องหรือผลกระทบของเลนส์ และ ข้อตผู้ชม เป็นข้อตที่แสดงถึงผู้ชมการแข่งขัน

2.4 โครงสร้างของระบบการจำแนกวีดิทัศน์กีฬาด้วยลักษณะเฉพาะวีดิทัศน์

เราแบ่งระบบการจำแนกวีดิทัศน์กีฬาด้วยลักษณะเฉพาะวีดิทัศน์ออกเป็นงานที่สำคัญ 3 ส่วนดังต่อไปนี้ [2]

2.4.1 การจำแนกข้อตเบื้องต้น

การจำแนกในส่วนนี้ถือว่ามีค่าความจำเป็นเนื่องจากวีดิทัศน์บางประเภทประกอบไปด้วยฉากโรงฉาย หรือ ฉากสัมภาษณ์ ซึ่งเป็นลักษณะเฉพาะของผู้สัมภาษณ์ ตัวอย่างหนึ่งของวีดิทัศน์ประเภทนี้ ได้แก่ วีดิทัศน์ของการแข่งขันกีฬาโอลิมปิกซึ่งบ่อยครั้งจะพบข้อตพิธีกรซึ่งเกิดจากการตัดต่อข้อตวีดิทัศน์จากสถานีโทรทัศน์ ดังนั้นจุดประสงค์ของส่วนนี้คือต้องการแยกข้อตวีดิทัศน์ที่เป็นฉากแข่งขันจากข้อตที่ไม่ใช่ฉากแข่งขัน ทั้งนี้สามารถนำแนวความคิดเชิงสถิติมาวิเคราะห์เนื้อหาของการมองเห็นที่เหมือนกันและลักษณะเฉพาะของการเคลื่อนที่ของข้อตผู้ประกาศ โดยไม่จำเป็นต้องสร้างแบบจำลองของข้อตเหล่านี้ โดยทั่วไปฉากโรงฉายจะถูกกำหนดด้วยจำนวนกล้องที่ใช้ มุมมองของกล้องที่มีอย่างจำกัด และ ลำดับของเนื้อหาภาพมักจะเกิดแบบรูปที่ซ้ำๆ กัน

2.4.2 การจำแนกลักษณะเฉพาะวัตถุกราฟฟิก

ในวีดิทัศน์กีฬาวัตถุกราฟฟิกเกิดขึ้นได้ทุกส่วนภายในเฟรม ส่วนใหญ่แล้ววัตถุกราฟฟิกปรากฏขึ้นบนเฟรมเป็นหนึ่งในสามหรือหนึ่งในสี่อยู่ที่ตำแหน่งด้านล่างของภาพ อัตราส่วนความกว้างและความยาวของบริเวณวัตถุกราฟฟิกมีการเปลี่ยนแปลงได้เสมอ ตัวอย่างเช่น วัตถุกราฟฟิกที่ใช้ในการแสดงชื่อของผู้แข่งขันอาจมีการวางตัวในแนวตั้งและแนวนอนก็ได้ บ่อยครั้งที่วัตถุกราฟฟิกจะปรากฏขึ้นและหายไปแบบค่อยเป็นค่อยไป นั่นคืออาจจะมีการปาดเข้าหรือปาดออกก็ได้ ดังนั้นส่วนนี้เป็นการตรวจหาวัตถุกราฟฟิกจากวีดิทัศน์กีฬาเพื่อดึงลักษณะเฉพาะวัตถุกราฟฟิกออกมา ตัวอย่างลักษณะเฉพาะวัตถุกราฟฟิก เช่น ข้อความอักษร เป็นต้น ในการตรวจหาวัตถุกราฟฟิกนี้จะใช้ลักษณะเฉพาะของขอบและพื้นผิวที่วางซ้อนกันเป็นวัตถุกราฟฟิก โดยที่ลักษณะเฉพาะเหล่านี้เป็นการแทนคุณสมบัติของภาพโดยรวมและต้องการวิเคราะห์ภาพที่มีขนาดใหญ่

2.4.3 การจำแนกลักษณะเฉพาะของการมองเห็น

จากที่กล่าวในหัวข้อ 2.3 ฉากการแข่งขัน เราสามารถแบ่งฉากการแข่งขันได้เป็นข้อตสนามแข่งขัน ข้อตนักกีฬา และ ข้อตผู้ชม ดังนั้นในขั้นตอนนี้เป็นการทำงานให้เป็นการแยกส่วนให้เป็นหมวดหมู่

ก่อน จากบทความที่ [2] แนะนำให้ใช้ลักษณะเฉพาะของขอบและรูปร่างสามารถช่วยในการแยกแยะความแตกต่างของข้อดีวิธีทัศน์ต่างๆ ได้

2.5 ลักษณะระดับล่างของวิทัศน์

ในส่วนนี้จะกล่าวถึงลักษณะเฉพาะระดับล่างของวิทัศน์ที่ได้จากระบบการดึงลักษณะเฉพาะเพื่อนำมาใช้ในการจำแนกประเภทกีฬา ประกอบด้วยลักษณะเฉพาะของสีในหัวข้อที่ 2.5.1 ลักษณะเฉพาะของขอบสีในหัวข้อที่ 2.5.2 และ ลักษณะเฉพาะของการเคลื่อนไหวในหัวข้อ 2.5.3

2.5.1 ลักษณะสี (Color Feature)

สิ่งที่มนุษย์สังเกตเห็นอย่างแรกคือสี ทำให้สีเป็นลักษณะที่โดดเด่น และ มีความสำคัญอย่างยิ่ง ปริภูมิสี (color space) แบ่งออกได้หลายรูปแบบ เช่น RGB HSI และ YUV เป็นต้น ความแตกต่างของปริภูมิสีที่ได้กล่าวนี้สำหรับ ปริภูมิ HSI และ YUV จะแยกองค์ประกอบของแสงสว่าง (luminance) และ ค่าสี (chrominance) ออกจากกัน ส่วนปริภูมิสี RGB จะแยกองค์ประกอบของสีเป็น องค์ประกอบสีแดง สีเขียว และ สีฟ้า ซึ่งในการจัดเก็บในรูปแบบของภาพส่วนใหญ่จัดเก็บในปริภูมิสี RGB โดยที่จุดภาพที่มีสีขนาด 24 บิตซึ่งในแต่ละองค์ประกอบสีจะมีระดับค่าสีเป็น 8 บิตซึ่งแบ่งได้เป็น 256 ระดับ (ค่าตั้งแต่ 0 ถึง 255) ดังนั้นความเป็นไปได้ของสีมากถึง 16.7 ล้านสี (256x256x256)

ลักษณะของสีที่ได้ศึกษา และ นำมาใช้งานในระบบค้นคืนวิทัศน์ประกอบไปด้วยฮิสโตแกรมสีรวม และ คอรีโลแกรมของสี

2.5.1.1 ฮิสโตแกรมสีรวม (Global Color Histogram) [26]

ฮิสโตแกรมสีเป็นการคำนวณหาความถี่ของสีภายในภาพ จากที่กล่าวถึงความเป็นไปได้ของสีที่จำนวนมากถึง 16.7 ล้านสี ดังนั้นเพื่อลดมิติ (Dimension) ของสีลงเพื่อใช้ในการสร้างดัชนีลักษณะสีจึงต้องลดระดับสีลงโดยการควอนไทซ์สีภายในภาพ อีกทั้งเพื่อลดความซับซ้อนของการคำนวณค่าฮิสโตแกรม โดยการแบ่งกลุ่มสีออกมาเป็น m ถังสี (Bin) ซึ่งมีเวกเตอร์ลักษณะเป็น (h_1, h_2, \dots, h_m) เมื่อในแต่ละ h_i แทนด้วยร้อยละของสีภายในภาพ

กำหนดให้ภาพ I มีขนาด $n_1 \times n_2$ จุดภาพ และ p เป็นจุดภาพที่อยู่ภายในภาพ I สามารถนิยามฮิสโตแกรมสี $h_{c_i}(I)$ เป็นความน่าจะเป็นที่จุดภาพ p มีสีเป็น c_i ซึ่งเขียนได้ดังสมการที่ (2.13)

$$h_{c_i}(I) = \text{probability}[p \in I_{c_i}] = \frac{\|I_{c_i}\|}{n_1 \times n_2} \quad (2.13)$$

โดยที่ $\|I_{c_i}\|$ เป็นจำนวนสีจุดภาพที่มีสี c_i ภายในภาพ I

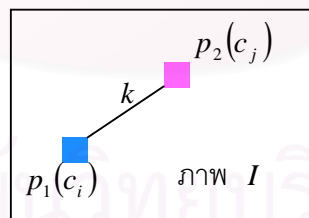
ดังนั้นจะได้เวกเตอร์แทนฮิสโตแกรมของสีของภาพ I ดังสมการที่ (2.14)

$$H(I) = (h_{c_1}, h_{c_2}, \dots, h_{c_m}) \quad (2.14)$$

ฮิสโตแกรมของสี เป็นลักษณะที่ใช้กันอย่างแพร่หลายในระบบค้นคืนภาพ เนื่องจากคำนวณได้ง่าย และ รวดเร็ว อีกทั้งกระบวนการไม่มีความซับซ้อน แต่ข้อเสียของลักษณะนี้คือละเลยลักษณะเชิงพื้นที่ของภาพ

2.5.1.2 คอร์ริโลแกรมของสี

คอร์ริโลแกรมของสี ซึ่งนิยามโดย [27] คือ กราฟหรือตารางที่แสดงสหสัมพันธ์เชิงพื้นที่ของคู่สีที่แตกต่างกันกับระยะห่างระหว่างคู่นั้น ซึ่งจะแตกต่างจากฮิสโตแกรมของสีที่สามารถอธิบายได้เฉพาะข้อมูลเกี่ยวกับการกระจายของสีภายในภาพเท่านั้น



รูปที่ 2.9 สหสัมพันธ์เชิงพื้นที่ระหว่างจุดภาพ p_1 และ p_2

ลักษณะเด่นของคอร์ริโลแกรมของสี [7] คือ

1. สามารถอธิบายสหสัมพันธ์เชิงพื้นที่ (spatial correlation) ระหว่างคู่สีต่าง ๆ ภายในภาพ ซึ่งเปลี่ยนแปลงไปตามระยะห่างระหว่างจุดภาพของคู่นั้น ๆ รวมถึงการกระจายตัวของสีภายในภาพได้
2. การคำนวณทำได้ง่าย

3. ขนาดของเวกเตอร์ลักษณะเฉพาะค่อนข้างเล็ก
4. สามารถใช้ในการค้นคืนภาพสำหรับระบบฐานข้อมูลที่มีขนาดใหญ่ได้อย่างมีประสิทธิภาพ
5. ทนทานต่อการเปลี่ยนแปลงของลักษณะและรูปร่างของวัตถุภายในภาพ ซึ่งเกิดจากการเปลี่ยนแปลงตำแหน่งและทิศทางในการมอง (viewing position) รวมถึงภาพที่มีการซูมจากกล้อง (camera zoom) และภาพที่มีสี, ความส่องสว่างและความเปรียบต่างเปลี่ยนแปลงไปจากเดิม

คอรีโลแกรมของสีระหว่างสี c_i กับ c_j ของภาพ I ที่ระยะห่างระหว่างจุดภาพของทั้งสองสีเท่ากับ k คือ ค่าความน่าจะเป็นที่จะพบจุดภาพที่มีสี c_j ที่ระยะห่างเท่ากับ k จากจุดภาพที่มีสี c_i ดังแสดงใน รูปที่ 2.9

การคำนวณคอรีโลแกรมของสี

กำหนดให้ I คือ ภาพสี ขนาด $n_1 \times n_2$ จุดภาพ $p(x, y)$ คือ จุดภาพที่พิกัด (x, y)
 m คือ จำนวนสีภายในภาพ หลังจากการควอนไทซ์สี โดยมีเซตของสีเป็น $\{c_1, \dots, c_m\}$
 $I(p)$ คือ ค่าสีของจุดภาพ $p = (x, y) \in I$
 $I_c \equiv \{p | I(p) = c\}$ ดังนั้น $p \in I_c$ จะสมมูลกับ $p \in I$ และ $I(p) = c$
 $[n]$ คือ เซตของ $\{1, 2, \dots, n\}$ และ $d \in [n]$
 p_1 คือ จุดภาพที่มีสี c_i
 p_2 คือ จุดภาพซึ่งอยู่ในภาพ I
 I_{c_j} คือ เซตของจุดภาพที่มีสี c_j ของภาพ I

ดังนั้นจะได้คอรีโลแกรมของสีของภาพ I สำหรับ i และ $j \in [m]$ และ $k \in [d]$ ดังสมการที่ (2.15)

$$\gamma_{c_i, c_j}^{(k)}(I) \equiv \Pr[p_2 \in I_{c_j} \mid p_1 - p_2 = k] \quad (2.15)$$

โดยที่ $\gamma_{c_i, c_j}^{(k)}(I)$ คือ เวกเตอร์แทนคอรีโลแกรมของสีของภาพ I
 k คือ ระยะห่างระหว่างจุดภาพ 2 จุดภาพใด ๆ

สำหรับระยะห่างระหว่างจุดภาพ 2 จุดภาพใด ๆ (k) คำนวณได้จาก L_∞ -norm โดยกำหนดให้จุดภาพ $p_1 = (x_1, y_1)$ และ $p_2 = (x_2, y_2)$ จะได้ว่าระยะห่างระหว่างจุดภาพทั้งสองมีค่าดังสมการที่ (2.16)

$$\|p_1 - p_2\| = \max\{|x_1 - x_2|, |y_1 - y_2|\} \quad (2.16)$$

ซึ่งการคำนวณด้วยวิธีนี้ เรียกว่า chessboard distance โดยที่ระยะ k เท่ากับ 1 สำหรับจุดภาพใด ๆ จะมีจำนวนจุดภาพรอบ ๆ จุดภาพนั้นเท่ากับ 8 จุดภาพ ดังนั้นที่ระยะ k สำหรับจุดภาพใด ๆ จะมีจำนวนจุดภาพรอบ ๆ เท่ากับ $8k$ จุดภาพ ดังพิจารณาได้จากรูปที่ 2.10 และรูปที่ 2.11

1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8
2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8
3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8
4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8
5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8
6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8
7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8
8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8

รูปที่ 2.10 ตัวอย่างการพิจารณาจุดภาพ (x_i, y_i) ที่ระยะ k เท่ากับ 1

1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8
2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8
3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8
4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8
5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8
6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8
7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8
8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8

รูปที่ 2.11 ตัวอย่างการพิจารณาจุดภาพ (x_i, y_i) ที่ระยะ k เท่ากับ 2

โดยที่ ■ คือ จุดภาพอ้างอิง

■ คือ จุดภาพที่ระยะ k รอบ ๆ จุดภาพอ้างอิง

จากรูปที่ 2.10 จะได้ว่าจำนวนจุดภาพที่ระยะ k เท่ากับ 1 รอบจุดภาพที่พิกัด (2,2) มีจำนวนเท่ากับ 8 จุดภาพ และจากรูปที่ 2.5 จำนวนจุดภาพที่ระยะ k เท่ากับ 2 รอบจุดภาพที่พิกัด (3,3) มีจำนวนเท่ากับ 16 จุดภาพ โดยในการคำนวณจะไม่สนใจจุดภาพที่อยู่นอกเหนือบริเวณภาพ ตัวอย่างเช่น ในรูปที่ 2.10 ค่าคอรีโลแกรมของสีระหว่างสีฟ้ากับสีฟ้า (คอรีโลแกรมระหว่างสีเดียวกันหรือออโตคอรีโลแกรมซึ่งจะอธิบายในหัวข้อถัดไป) ของจุดภาพที่พิกัด (1,1) ที่ระยะ k เท่ากับ 1 จะเท่ากับ $2/8$ และสำหรับจุดภาพที่พิกัด (1,2) จะมีค่าเท่ากับ $4/8$ เป็นต้น

2.5.1.3 ออโตคอรีโลแกรม

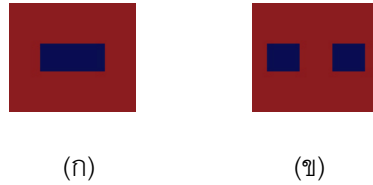
ออโตคอรีโลแกรม เป็นเซตย่อย (subset) ของคอรีโลแกรมของสี โดยออโตคอรีโลแกรมของภาพ I ที่ระยะ k คือ ค่าความน่าจะเป็นที่จะพบจุดภาพที่มีสีเดียวกันที่ระยะห่างระหว่างจุดภาพเท่ากับ k ซึ่งคำนวณได้จากสมการที่ (2.17)

$$\gamma_{c_i, c_i}^{(k)}(I) \equiv \Pr[p_2 \in I_{c_i} \mid p_1 - p_2 = k] \quad (2.17)$$

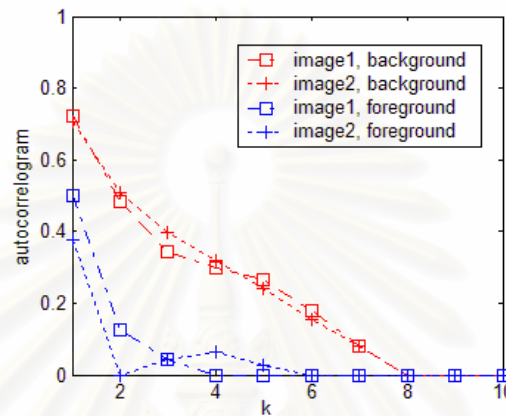
โดยกำหนดให้ $\gamma_{c_i, c_i}^{(k)}(I)$ คือ เวกเตอร์แทนออโตคอรีโลแกรมของภาพ I สำหรับ $i \in [m]$ และ $k \in [d]$ ดังสมการที่ (2.18)

$$\gamma_{C_i, C_i}^{(k)}(I) = \begin{bmatrix} \gamma_{c_1, c_1}^{(1)} & \gamma_{c_1, c_1}^{(2)} & \cdots & \gamma_{c_1, c_1}^{(k)} \\ \gamma_{c_2, c_2}^{(1)} & \gamma_{c_2, c_2}^{(2)} & \cdots & \gamma_{c_2, c_2}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{c_m, c_m}^{(1)} & \gamma_{c_m, c_m}^{(2)} & \cdots & \gamma_{c_m, c_m}^{(k)} \end{bmatrix} \quad (2.18)$$

ข้อดีของออโตคอรีโลแกรมเมื่อเทียบกับคอรีโลแกรมของสี คือ มีขนาดของเวกเตอร์ที่เล็กกว่า และความซับซ้อนในการคำนวณน้อยกว่าทำให้สามารถคำนวณได้อย่างรวดเร็ว ซึ่งความซับซ้อนในการคำนวณออโตคอรีโลแกรมของภาพจะแปรตามจำนวนสีภายในภาพ และเซตของระยะห่างระหว่างจุดภาพ 2 จุดภาพใด ๆ (k) ซึ่งในงานวิจัย [27] ใช้ค่า $k = \{1, 3, 5, 7, 9\}$ ข้ออีกประการหนึ่งของออโตคอรีโลแกรมเมื่อเทียบกับฮิสโตแกรมของสี พิจารณาได้จากภาพตัวอย่าง 2 ภาพในรูปที่ 2.12 ซึ่งทั้งสองภาพมีฮิสโตแกรมของสีที่เหมือนกัน กล่าวคือ จำนวนจุดภาพที่มีสีแดงและสีน้ำเงินของทั้งสองภาพนั้นมีจำนวนที่เท่ากัน ดังนั้นในกรณีนี้จึงไม่สามารถแยกแยะความแตกต่างระหว่างภาพ 2 ภาพนี้ได้โดยการใช้ฮิสโตแกรมของสี



รูปที่ 2.12 ภาพตัวอย่างที่มีฮิสโตแกรมของสีเหมือนกัน
(ก) ภาพที่ 1 (image 1) (ข) ภาพที่ 2 (image 2)



รูปที่ 2.13 ออโตคอรีโลแกรมของภาพตัวอย่างในรูปที่ 2.12

จากรูปที่ 2.13 จะสังเกตเห็นว่าค่าออโตคอรีโลแกรมของแต่ละสีที่ระยะ k ค่าต่างๆ ของทั้งสองภาพมีค่าที่แตกต่างกัน ดังนั้นจึงทำให้สามารถแยกแยะความแตกต่างระหว่างภาพทั้งสองภาพนี้ได้ เนื่องจากออโตคอรีโลแกรม สามารถอธิบายสหสัมพันธ์เชิงพื้นที่ระหว่างคู่สีที่เหมือนกันภายในภาพได้ ทำให้ออโตคอรีโลแกรมมีความทนทานต่อการเปลี่ยนแปลงของภาพมากกว่าวิธีอื่นๆ ดังนั้นจึงสามารถสร้างดัชนีและค้นคืนข้อมูลภาพ ซึ่งอยู่ในสถานที่เดียวกันแต่ต่างมุมมอง รวมไปถึงภาพที่มีการซูมจากกล้อง, ภาพที่มีสี, ความส่องสว่างที่การเปลี่ยนแปลงได้อย่างมีประสิทธิภาพ

2.5.2 ลักษณะการเคลื่อนที่ของกล้อง

ลักษณะทางเวลาเป็นลักษณะที่อธิบายการเปลี่ยนแปลงของเนื้อหาในทางเวลา ซึ่งถือเป็นคุณสมบัติของเนื้อหา หรือ เรื่องราวที่เกิดขึ้นในวิดีโอ เช่น การเดิน การวิ่ง การกระโดด หรือ การบิน เป็นต้น หลักการพื้นฐานที่สำคัญในหาลักษณะการเคลื่อนที่คือ การประมาณการเคลื่อนที่ (Motion Estimation) เป็นการประมาณค่าเวกเตอร์การเคลื่อนที่ในแนว 2 มิติซึ่งเรียกว่าเวกเตอร์การไหลเชิงแสง (Optical Flow) เป็นค่าที่บอกถึงความเร็วในการเคลื่อนที่ที่สังเกตได้ เวกเตอร์การไหลเชิงแสงนี้ไม่เพียงบอกลักษณะการเคลื่อนที่ของวัตถุเท่านั้นทั้งยังบ่งบอกถึงการ

เคลื่อนที่ของกล้องได้อีกด้วย โดยที่เทคนิคการประมาณการเคลื่อนที่ส่วนใหญ่ได้จากการเปลี่ยนแปลงไปทางเวลาของค่าความส่องสว่าง หรือ ค่าสีดังสมการที่ (2.19)

$$\psi(x + d_x, y + d_y, t + d_t) = \psi(x, y, t) \quad (2.19)$$

เมื่อ $\psi(x, y, t)$ เป็นค่าความส่องสว่าง โดยที่ภาพที่จุดภาพ (x, y) ที่เวลา t ที่มีการเคลื่อนที่ไป ณ จุดภาพ $(x + d_x, y + d_y)$ ที่เวลาที่ $t + d_t$ โดยที่ค่าความส่องสว่างของจุดภาพจุดเดียวกันไม่มีการเปลี่ยนแปลงค่า เมื่อใช้ Taylor's expansion เมื่อ d_x , d_y และ d_t มีขนาดเล็กแล้วจะได้ว่า

$$\psi(x + d_x, y + d_y, t + d_t) = \psi(x, y, t) + \frac{\partial \psi}{\partial x} d_x + \frac{\partial \psi}{\partial y} d_y + \frac{\partial \psi}{\partial t} d_t \quad (2.20)$$

จากสมการที่ (2.19) และ (2.20) จะได้ว่า

$$\frac{\partial \psi}{\partial x} d_x + \frac{\partial \psi}{\partial y} d_y + \frac{\partial \psi}{\partial t} d_t = 0 \quad (2.21)$$

สมการที่ (2.21) เขียนอยู่เทอมของเวกเตอร์การเคลื่อนที่ (Motion Vector) (d_x, d_y) ถ้าหารทั้งข้างสองด้วย d_t แล้วจะได้ว่า

$$\frac{\partial \psi}{\partial x} v_x + \frac{\partial \psi}{\partial y} v_y + \frac{\partial \psi}{\partial t} = 0 \quad \text{or} \quad \nabla \psi^T + \frac{\partial \psi}{\partial t} = 0 \quad (2.22)$$

โดยที่ (v_x, v_y) แทนด้วยเวกเตอร์ความเร็ว (Velocity vector) หรือ เวกเตอร์การ

ไหล (Flow vector) และ $\nabla \psi^T = \left[\frac{\partial \psi}{\partial x}, \frac{\partial \psi}{\partial y} \right]^T$ สมการที่ (2.22) เรียกว่า สมการไหลเชิงแสง

(optical flow equation) จากสมการที่ (2.23) พบว่ามีสองพารามิเตอร์ที่ไม่ทราบค่า (v_x, v_y) แต่มีเพียง 1 สมการเท่านั้นจึงจำเป็นต้องให้เทคนิคการประมาณการเข้าช่วยเพื่อหาพารามิเตอร์ทั้งสองที่ต้องการ

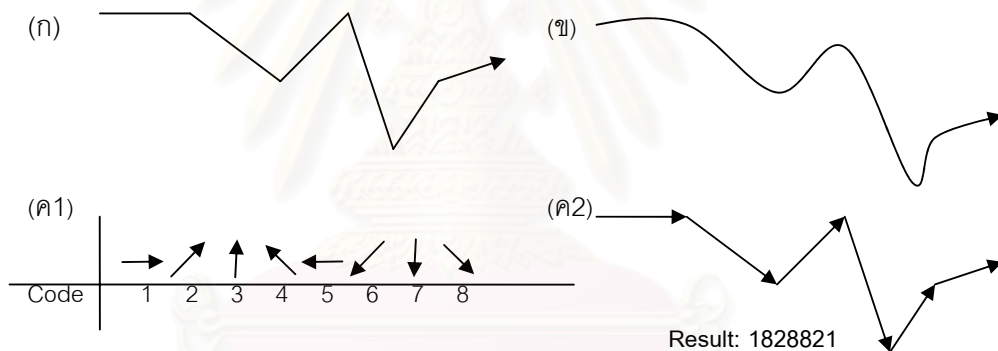
เทคนิคในการตั้งลักษณะทางเวลาส่วนใหญ่ล้วนแต่เริ่มต้นด้วยหาค่าเวกเตอร์การเคลื่อนที่ซึ่งได้จากการประมาณการเคลื่อนที่ได้เทคนิคต่างๆ ที่ได้เสนอไว้ในบทความ [28] เมื่อได้เวกเตอร์การเคลื่อนที่แล้วจึงค่อยประมาณลักษณะทางเวลาอีกต่อไปซึ่งจำแนกเทคนิคการตั้งคุณลักษณะทางเวลาเป็น 3 แบบดังต่อไปนี้คือ

- แนวคิดพื้นฐานลักษณะ (Feature Based Approach)

เทคนิคนี้เริ่มด้วยการคำนวณพารามิเตอร์การเคลื่อนที่แล้วประมาณการโดยใช้แบบจำลองการเคลื่อนที่ เช่น translation model, affine Model หรือ bilinear Model เพื่อบอกลักษณะการเคลื่อนที่เช่นบทความ [29] และ [30] แล้วนำพารามิเตอร์ที่ได้ดังกล่าวมาทำเป็นดัชนีวีดิทัศน์

- แนวความคิดพื้นฐานแนวโน้มการเคลื่อนที่ (Trajectory Based Approach)

ข้อมูลวีดิทัศน์จะประกอบด้วยวัตถุวีดิทัศน์เดี่ยว หรือ วัตถุวีดิทัศน์หลายวัตถุที่มีการเคลื่อนที่ภายในข้อมูลวีดิทัศน์ ดังนั้นแนวการเคลื่อนที่ของวัตถุวีดิทัศน์จึงเป็นหนึ่งที่สามารถนำมาใช้เป็นดัชนีในการค้นคืนได้ เช่นบทความ [31] ซึ่งสามารถสรุปลักษณะตามแนวการเคลื่อนที่ได้ดังรูปที่ 2.14



รูปที่ 2.14 ลักษณะการเคลื่อนที่ของวัตถุแบบต่าง

(ก) แนวการเคลื่อนที่ (motion tractory) (ข) การแทนแนวการเคลื่อนที่ B-Spline

(ค1) รูปแบบของรหัสลูกโซ่ และ (ค2) รหัสลูกโซ่ที่แทนแนวทางการเคลื่อนที่

จากแนวความคิดที่ว่ากีฬาต่างชนิดมีแบบรูปของการเคลื่อนที่แตกต่างกัน ตัวอย่างเช่น ในกีฬาบาสเกตบอล ลักษณะเฉพาะการเคลื่อนที่ของกล้องในการถ่ายและขวจะ มีการเปลี่ยนแปลงอย่างทันทีและยังรวดเร็วสำหรับการบุกและการป้องกันซึ่งบ่งบอกว่าความถี่ในการถ่ายมีมาก แต่ในกีฬาฟุตบอล ลักษณะการเคลื่อนที่ของกล้องในการถ่ายและขวจะให้ความถี่ของการเปลี่ยนแปลงจากการถ่ายที่ให้ค่าน้อย [17] ลักษณะเฉพาะของการเคลื่อนที่ได้จากการประมาณการลักษณะการเคลื่อนที่โดยใช้พื้นฐานบล็อก (block based motion estimation)

ซึ่งเป็นการหาผลต่างของเฟรมในระดับบล็อกในการชดเชยการเคลื่อนที่เฉลี่ย (mean motion compensated block frame difference) [32] ดังสมการที่ (2.23)

$$Motion_{B_k} = \min_{(u,v)} \sum_{(x,y) \in B_k} (F_i(x,y) - F_{i-1}(x+u, y+v)) \quad (2.23)$$

เมื่อ B_k แทนบล็อกๆหนึ่งภายในภาพ F_i แทนเฟรมตำแหน่งที่ i ของวิดีโอทัศน์ และ (u,v) เป็นระยะค้นหา (search range) ของเวกเตอร์การเคลื่อนที่ซึ่งเหมือนกับการเข้ารหัสใน MPEG ซึ่งผลที่ได้จากสมการที่ (2.23) ในแต่ละบล็อกจะประกอบเป็นสนามการเคลื่อนที่ (Motion Field)

หลังจากได้เวกเตอร์การเคลื่อนที่ในแต่ละบล็อก (u,v) ซึ่งได้จากสมการที่ (2.23) เราสามารถคำนวณขนาดเวกเตอร์การเคลื่อนที่ (magnitude of motion vector) ในแต่ละบล็อก จากสมการที่ (2.24)

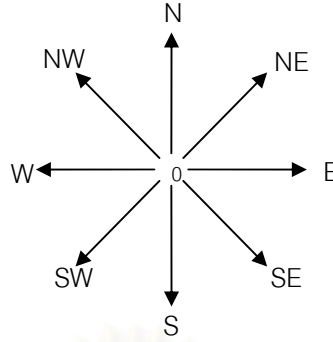
$$Mag_{B_k} = \sqrt{u_k^2 + v_k^2} \quad (2.24)$$

โดยที่ (u_k, v_k) เป็นเวกเตอร์การเคลื่อนที่ในบล็อกที่ k เรากำหนดให้เซตสัญลักษณ์ของทิศทางการเคลื่อนที่ประกอบด้วยสมาชิกจำนวน 10 สมาชิก เป็น

$$S = \{S_{-1}, S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\} \quad (2.25)$$

โดยที่เป็นสัญลักษณ์การแทนดังต่อไปนี้ S_{-1} แทนไม่ทราบการเคลื่อนที่ S_0 แทนไม่มีการเคลื่อนที่ S_1 แทนการเคลื่อนที่ในทิศตะวันออก (E) S_2 แทนการเคลื่อนที่ในทิศตะวันออกเฉียงเหนือ (NE) S_3 แทนการเคลื่อนที่ในทิศเหนือ (N) S_4 แทนการเคลื่อนที่ในทิศตะวันตกเฉียงเหนือ (NW) S_5 แทนการเคลื่อนที่ในทิศตะวันตก (W) S_6 แทนการเคลื่อนที่ในทิศตะวันตกเฉียงใต้ (SW) S_7 แทนการเคลื่อนที่ในทิศใต้ (S) และ S_8 แทนการเคลื่อนที่ในทิศตะวันออกเฉียงใต้ (SE) ดังรูปที่ 2.15

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.15 การกำหนดทิศทางหลักของเวกเตอร์การเคลื่อนที่

ในแต่ละบล็อก B_k ของเฟรมจะมีเวกเตอร์การเคลื่อนที่เป็น $\vec{p} = (u, v)$ ดังนั้นสามารถคำนวณมุมระหว่าง \vec{p} กับ \vec{d} ซึ่งเป็นค่าเวกเตอร์ในแนวแกนหลักในรูปที่ 2.15 ได้จากสมการที่ (2.26)

$$\theta_i = \arccos \left(\frac{\vec{p} \cdot \vec{d}_i}{\|\vec{p}\| \|\vec{d}_i\|} \right) \quad (2.26)$$

เมื่อ θ_i เป็น มุมระหว่างระหว่าง \vec{p} กับ \vec{d} ซึ่งหมายถึงระยะทางระหว่างเวกเตอร์การเคลื่อนที่กับเวกเตอร์สัญลักษณ์ที่กำหนดขึ้นดังรูปที่ 2.15 จากนั้นสามารถกำหนดสัญลักษณ์ให้กับแต่ละบล็อก B_k ได้จากสมการที่ (2.27)

$$S_i = \arg \min(\theta_i), \quad i = 1, 2, \dots, 8 \quad (2.27)$$

ถ้าขนาดเวกเตอร์การเคลื่อนที่ในแต่ละบล็อกมีค่าน้อย เรากำหนดสัญลักษณ์ให้กับบล็อกเป็น S_0 หลังจากที่กำหนดลักษณะให้กับบล็อกแล้วเราจะกำหนดสัญลักษณ์ในระดับเฟรมได้จากฮิสโตแกรมของสัญลักษณ์ที่กำหนดในระดับบล็อก โดยกำหนดให้ในแต่ละเฟรมมีจำนวนบล็อกเป็น N บล็อกของเฟรม F และฮิสโตแกรมของสัญลักษณ์แทนด้วย $h(S_j)$ ดังนั้นสามารถกำหนดสัญลักษณ์ให้กับเฟรมได้จากสมการที่ (2.28)

$$h(S_i) = \arg \max_j (h(S_j)) \quad j = 0, 1, \dots, 8 \quad (2.28)$$

2.5.3 ลักษณะขอบ

ลักษณะขอบถือว่าเป็นลักษณะอีกประเภทหนึ่งของภาพที่นิยมนำมาใช้ในงานทางด้าน การสร้างดัชนี และการค้นคืนภาพ ซึ่งลักษณะของขอบหมายถึงรูปร่างของวัตถุภายในภาพ [27] ทั้งนี้สามารถหาลักษณะของขอบได้จากการประมวลก่อนหน้า (preprocessing) หรือตัวกรอง (filter) โดยภาพที่ได้ที่มาจากตัวกรองถูกเรียกว่าภาพขอบวัตถุ นั่นคือ กลุ่มของจุดภาพที่

นำมาวางต่อๆ กันบนเส้นแบ่งเขตระหว่างพื้นที่สองส่วนของวัตถุภายในภาพ ซึ่งจะช่วยอธิบายถึงรูปร่างและลักษณะ รวมถึงขนาดของวัตถุภายในภาพ ขอบของวัตถุนิยามด้วยการเปลี่ยนแปลงความไม่ต่อเนื่องหรือความไม่สม่ำเสมอ (discontinuity) ของระดับความเข้มสีของภาพระดับเทา (gray levels) ดังนั้นในการตรวจหาขอบของวัตถุจึงทำได้โดยอาศัยตัวดำเนินการอนุพันธ์ (derivative operator) โดยที่อนุพันธ์ของฟังก์ชันก็คืออัตราการเปลี่ยนแปลงของฟังก์ชัน ดังนั้นบริเวณขอบของวัตถุควรเป็นบริเวณที่มีอัตราส่วนการเปลี่ยนแปลงของระดับเทามาก ส่วนบริเวณที่มีระดับสีเทาสม่ำเสมอก็จะมี การเปลี่ยนแปลงของระดับความเข้มสีเทาน้อย ซึ่งภาพที่เกิดขึ้นจริงโดยส่วนใหญ่จะมีการเปลี่ยนแปลงของระดับความเข้มสีเทาอย่างต่อเนื่อง ซึ่งตรงกันข้ามกับกับในทางอุดมคติซึ่งมีการเปลี่ยนแปลงของระดับความเข้มของสีเทาอย่างทันที ณ ตำแหน่งขอบ

2.5.3.1 การตรวจหาขอบวัตถุภายในภาพ

การตรวจหาขอบวัตถุ คือ กระบวนการตรวจหาตำแหน่งหรือที่ตั้งของจุดภาพซึ่งขอบ (edge pixels) ของวัตถุภายในภาพ โดยที่ส่วนใหญ่บริเวณขอบของวัตถุจะเป็นบริเวณที่มีความถี่สูง (high frequencies) ในทางทฤษฎีเราสามารถตรวจหาขอบวัตถุได้ 2 วิธี คือ

- การกรองด้วยตัวกรองผ่านความถี่สูง (high pass frequency filter) ในโดเมนความถี่ (frequency domain)
- การคอนโวลูชัน (convolution) ภาพด้วยเคอร์เนล (kernel) ในโดเมนเชิงพื้นที่ (spatial domain)

การตรวจหาขอบวัตถุในโดเมนเชิงพื้นที่โดยอาศัยการทำอนุพันธ์แบ่งได้เป็น 2 วิธี ประกอบด้วย อนุพันธ์อันดับหนึ่ง (first order derivative) นำมาใช้ในการคำนวณอัตราการเปลี่ยนแปลงของระดับค่าสี เพื่อตรวจหาตำแหน่งขอบของวัตถุภายในภาพ โดยที่ตำแหน่งขอบจะมีอัตราการเปลี่ยนแปลงของระดับมากกว่าบริเวณอื่นๆ และ อนุพันธ์อันดับสอง (second order derivative) ใช้ในการตรวจหาขอบของวัตถุ โดยจุดภาพของขอบจะอยู่ที่ตำแหน่งซึ่งให้อัตราการเปลี่ยนแปลงของระดับสีที่คงที่ค่าหนึ่ง (มากกว่าศูนย์) นั่นคือจุดผ่านศูนย์ของอนุพันธ์อันดับสอง

เนื่องจากภาพมีลักษณะเป็น 2 มิติ ดังนั้นในการตรวจหาขอบวัตถุจึงต้องอาศัยตัวดำเนินการที่เรียกว่า “เกรเดียนต์ (gradient operation)” ซึ่งนิยามดังสมการที่ (2.29)

$$\nabla f(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} \end{bmatrix} \quad (2.29)$$

โดยกำหนดให้ $f(x, y)$ เป็นฟังก์ชันของภาพ และ G_x และ G_y เป็น เกรเดียนต์แนวนอน และ แนวตั้ง ตามลำดับ

ขนาดของเกรเดียนต์ $|G|$ คำนวณได้จากสมการที่ (2.30)

$$|G| = [G_x^2 + G_y^2]^{1/2} \quad (2.30)$$

โดยทิศทางการตรวจหาขอบวัตถุ $\alpha(x, y)$ คำนวณได้จากสมการที่ (2.31)

$$\alpha(x, y) = \tan^{-1} \left(\frac{G_y}{G_x} \right) \quad (2.31)$$

เนื่องจากภาพเป็นฟังก์ชันที่ไม่ต่อเนื่อง (discrete function) ดังนั้นแทนการทำอนุพันธ์ด้วยการ คำนวณผลต่าง โดยค่า G_x และ G_y ในสมการที่ (2.29) สามารถคำนวณได้จาก F_x และ F_y ดัง สมการที่ (2.32) และ (2.33) ตามลำดับ

$$F_x = f(x, y) - f(x-1, y) \quad (2.32)$$

$$F_y = f(x, y) - f(x, y-1) \quad (2.33)$$

โดยที่ขนาด $|F|$ และทิศทางของเกรเดียนต์ $\alpha_F(x, y)$ สามารถคำนวณได้จากสมการที่ (2.34) และ (2.35) ตามลำดับ

$$|F| = [F_x^2 + F_y^2]^{1/2} \quad (2.34)$$

$$\alpha_F(x, y) = \tan^{-1} \left(\frac{F_y}{F_x} \right) \quad (2.35)$$

ยกตัวอย่างการคำนวณค่า G_x และ G_y ในสมการที่ (2.30) ของจุดภาพ z_s ได้จากสมการที่ (2.36) และ (2.37) ตามลำดับ [33] โดยกำหนดค่าของจุดภาพ ตัวพรางสำหรับแนวแกน x และ ตัวพรางสำหรับแนวแกน y ดังนี้

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

รูปที่ 2.16 ค่าของจุดภาพ

-1	0
0	1

รูปที่ 2.17 ตัว
พรางสำหรับ
แนวแกน x

0	-1
1	0

รูปที่ 2.18 ตัว
พรางสำหรับ
แนวแกน y

$$G_x = (z_9 - z_5) \quad (2.36)$$

$$G_y = (z_8 - z_6) \quad (2.37)$$

ผลรวมสัมประสิทธิ์ตัวพรางจะส่งผลกระทบต่อระดับความส่องสว่างตลอดทั้งภาพ ซึ่งโดยทั่วไปผลรวมของค่าสัมประสิทธิ์ตัวพรางจะมีค่าเท่ากับ 1 แต่ถ้าผลรวมของค่าสัมประสิทธิ์ตัวพรางมีค่าน้อยกว่าหรือเท่ากับศูนย์ ภาพจะมีความมืดเพิ่มมากขึ้น ซึ่งจะพบได้ในการตรวจหาขอบวัตถุ ซึ่งในวิทยานิพนธ์นี้เลือกใช้ตัวพรางแบบ Sobel เนื่องจากใช้ขนาดตัวพรางที่มีขนาดใหญ่ จึงช่วยลดผลกระทบจากสัญญาณรบกวนที่ดี

2.5.3.2 ตัวพรางขอบวัตถุแบบ Sobel

การตรวจหาขอบวัตถุของภาพด้วยวิธีนี้สามารถคำนวณค่า G_x และ G_y ได้จากสมการที่ (2.38) และ (2.39) ตามลำดับ

$$G_x = I \square M_x \quad (2.38)$$

$$G_y = I \square M_y \quad (2.39)$$

โดยที่ M_x คือ ตัวพรางสำหรับการหาขอบวัตถุแนวอนหรือแนว 0 องศา ดังนี้

$$M_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \text{ หรือ } \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

M_y คือ ตัวพรางสำหรับการหาขอบวัตถุแนวตั้งหรือแนว 90 องศา ดังนี้

$$M_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ หรือ } \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

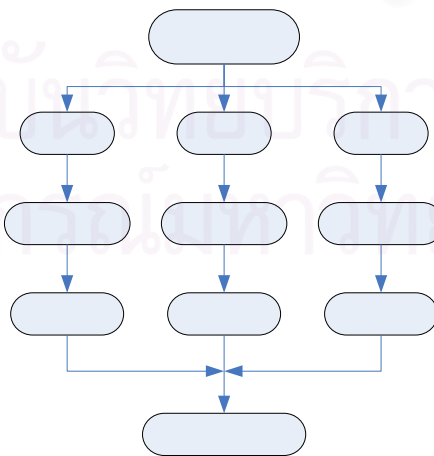
G_x คือเกรเดียนต์ของภาพ I ในแนวนอนหรือแนว 0 องศา

G_y คือเกรเดียนต์ของภาพ I ในแนวตั้งหรือแนว 90 องศา

หมายเหตุ : สัญลักษณ์ \square แทนด้วยการคำนวณภาพเกรเดียนต์

2.5.3.3 การตรวจหาขอบสี

โดยทั่วไปการตรวจหาขอบวัตถุจากการหาเกรเดียนต์จากองค์ประกอบความส่องสว่าง (luminance) ของภาพเท่านั้น ในกรณีที่ภาพมีขอบวัตถุภาพในองค์ประกอบความส่องสว่างที่ลักษณะเหมือนกัน แต่สามารถแยกแยะได้โดยการหาขอบภาพจากองค์ประกอบของสีอื่นๆ จึงมีการนำเสนอการตรวจหาขอบสี ซึ่งเป็นการตรวจหาขอบในแต่ละองค์ประกอบของปริภูมิสี เช่นเดียวกับการตรวจหาขอบโดยทั่วไป เทคนิคการหาขอบสีของวัตถุมีขั้นตอนดังรูปที่ 2.19 [34]



รูปที่ 2.19 แผนภูมิสายงานการตรวจหาขอบสี

ถึงแม้ว่าปริภูมิสีมีหลายแบบ เช่น RGB HSV และ YUV เป็นต้น แต่ปริภูมิสีที่เลือกใช้ในสำหรับการตรวจหาขอบสีของวัตถุในวิทยานิพนธ์นี้คือ YUV เนื่องจากปริภูมิสี YUV แยกองค์ประกอบของสี (U และ V) ออกจากองค์ประกอบความส่องสว่าง ซึ่งการตรวจหาขอบสีได้ในแต่ละองค์ประกอบของสีได้จากสมการที่ (2.40) และ (2.41) ตามลำดับ

$$G_x^i = I^i \square M_x \quad i \in \{Y, U, V\} \quad (2.40)$$

$$G_y^i = I^i \square M_y \quad i \in \{Y, U, V\} \quad (2.41)$$

2.5.3.4 ความหมายของขอบและการสร้างโครงสร้างขอบแบบรูป

เนื่องจากลักษณะเฉพาะขอบสามารถนำไปใช้แทนรูปร่างวัตถุได้ ทำให้สามารถนำมาสร้างลักษณะเชิงความหมายได้โดยกำหนดความหมายให้กับลักษณะเฉพาะขอบ การนำลักษณะเฉพาะขอบมากำหนดลักษณะเชิงความหมายมีแนวคิดดังต่อไปนี้ [21]

- วัตถุในทุกภาพสามารถอธิบายได้โดยลักษณะเฉพาะของขอบ
- วัตถุในฉากหน้าของภาพมีความสัมพันธ์กับลักษณะเชิงความหมาย
- การวางตัวของขอบ (edge orientation) ที่เกิดขึ้นสามารถสังเคราะห์ได้ใช้ 5 ทิศทางขอบคือ แนวตั้ง (|) แนวนอน (-) แนว 45 องศา (/) แนว 135 องศา (\) และ ไม่มีขอบ

จากสมมุติฐานนี้ ภาพจะถูกแบ่งออกเป็นบล็อก ในแต่ละบล็อกถูกกำหนดการวางตัวของขอบ โดยใช้ฮิสโตแกรมของความน่าจะเป็นที่เกิดขึ้นในแต่ละทิศทางที่ให้ค่าฮิสโตแกรมมากที่สุด เมื่อนำการหาขนาดและทิศทางเกรเดียนต์จากสมการที่ (2.34) และ (2.35) มาใช้ โดยในแต่ละจุดภาพมีเกรเดียนต์เวกเตอร์ เป็น $G(x, y) = (G_x, G_y)$ ซึ่งสามารถคำนวณมุมระหว่าง G ในแต่ละจุดภาพกับ เวกเตอร์หลัก D_i ของ 4 ทิศทางขอบคือ เวกเตอร์ในแนวตั้ง (|) แนวนอน (-) แนว 45 องศา (/) และ แนว 135 องศา (\) ดังสมการที่ (2.42)

$$\theta_i = \arccos \left(\frac{|G \cdot D_i|}{\|G\| \|D_i\|} \right) \quad (2.42)$$

ดังนั้นเราสามารถกำหนดการวางตัวของขอบในแต่ละบล็อกได้จากสมการที่ (2.43)

$$E^{B_k}_i = \arg \min (\theta_i), \quad i = -, /, \backslash \quad (2.43)$$

สำหรับการวางตัวแบบไม่มีขอบ (null) หาได้จากขนาดเกรเดียนต์น้อยกว่าขีดเริ่มเปลี่ยนค่าหนึ่ง ดังนั้นเวกเตอร์ลักษณะเฉพาะขอบ E แทนได้ดังสมการที่ (2.44)

$$E = (E^{B_1} E^{B_2} \dots E^{B_N}) \quad E^{B_k} \in \{-, |, /, \backslash, null\} \quad (2.44)$$

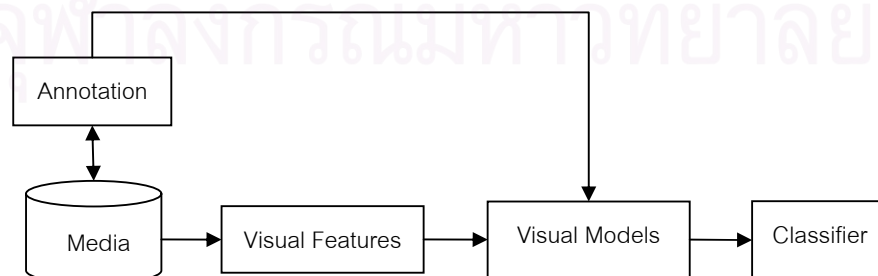
โดยที่ ภาพแบ่งได้เป็นจำนวน N บล็อก ซึ่งในวิทยานิพนธ์นี้ลักษณะของขอบสีเข้าร่วมดังนั้นจากสมการที่ (2.44) เรียงเวกเตอร์ในแต่ละองค์ประกอบของปริภูมิสี (edge color) ได้ดังสมการที่ (2.45)

$$edgecolor = [E^Y E^U E^V] \quad (2.45)$$

2.6 การจำแนกประเภทวิดิทัศน์โดยใช้การรู้จำรูปแบบ

ประเภทของวิดิทัศน์ถือว่าเป็นลักษณะเชิงความหมายแบบหนึ่ง ซึ่งลักษณะเชิงความหมายคือลักษณะที่มนุษย์รับรู้และเข้าใจ เช่น ภาพหนึ่งมีรถเป็นองค์ประกอบของภาพ มนุษย์จะตีความหมายว่าเป็นภาพของรถ เป็นต้น ดังนั้นเพื่อดึงลักษณะเชิงความหมายออกมา จำเป็นต้องใช้เทคนิคที่ให้การเรียนรู้โดยเครื่องจักร (machine learning) หรือ การทำฉลากเชิงความหมาย (semantic labeling) ซึ่งแนวความคิดเชิงความหมายนิยมไว้ในบทความ [36] อยู่ในโครงสร้างของอะตอมที่ประกอบไปด้วย วัตถุ (object) ฉาก (scene) และ เหตุการณ์ (event)

ปัญหาของลักษณะเชิงความหมายประการแรกคือการหาลักษณะระดับล่างที่เหมาะสมเพื่อที่กำหนดฉลากให้กับภาพหรือวิดิทัศน์ จากนั้นหาแบบจำลองที่เหมาะสมกับลักษณะเชิงความหมาย ประการที่ 2 คือแนวความคิดเชิงความหมายจะต้องเชื่อมโยงกับลักษณะเชิงความหมายอื่นๆ เพื่อสร้างลักษณะเชิงความหมายในขั้นถัดไป ประการสุดท้ายการรวมลักษณะทั้งระดับบน และ ระดับล่างให้มีแนวความคิดที่มีความหมาย ซึ่งปัญหาที่สำคัญที่สุดคือการหาลักษณะระดับล่างที่กำหนดฉลากให้กับภาพ หรือ วิดิทัศน์



รูปที่ 2.20 แผนภาพการวิเคราะห์แนวคิดเชิงความหมาย

การวิเคราะห์แนวคิดเชิงความหมายดังรูปที่ 2.20 ประกอบด้วย ส่วนการให้คำประกอบ (annotation) เป็นการกำหนดความหมายที่ต้องการวิเคราะห์ให้ทำการเรียนรู้ซึ่งจะต้องกำหนดขึ้นโดยมนุษย์ สำหรับลำดับวิธีทัศน์จะทำในระดับข้อวิธีทัศน์ ส่วนลักษณะทางการมองเห็นเป็นลักษณะระดับล่างที่ได้จากการประมวลภาพและวิธีทัศน์ เมื่อได้ลักษณะระดับล่างแล้วจึงนำไปหาแบบจำลองที่ได้กำหนดขึ้นในส่วนการให้คำประกอบ

ลักษณะเชิงความหมายสร้างได้จากเทคนิคของการรู้จำแบบ (pattern recognition) ซึ่งเขียนอยู่ในรูปแบบของความน่าจะเป็นของคลาส ω_j (ในที่นี้หมายถึงประเภทของวิธีทัศน์ที่ต้องการรู้จำ) ที่เกิดขึ้นจากค่าเวกเตอร์ลักษณะระดับล่างที่วัด \mathbf{x} เป็นเวกเตอร์สุ่ม (random vector) ดังสมการที่ (2.46)

$$P(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j) P(\omega_j)}{p(\mathbf{x})} \quad (2.46)$$

โดยที่

$$p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x} | \omega_j) P(\omega_j) \quad (2.47)$$

เมื่อ $p(\mathbf{x} | \omega_j)$ เป็นลักษณะเชิงความหมายที่สร้างขึ้นโดยใช้แนวคิดแบบจำลองเชิงความหมายมี 2 แนวความคิด คือ แนวความคิดแบบจำลองเชิงสถิติ และ แนวความคิดแบบจำลองดิสคริมิแนนต์

- **แนวคิดแบบจำลองเชิงสถิติ (Statistical Modeling Approach)**

ในแนวความคิดแบบจำลองเชิงสถิติได้ตั้งสมมติฐาน ให้ลักษณะเชิงความหมายมีการแจกแจงแบบเกาส์ ดังสมการที่ (2.48)

$$p(\mathbf{x} | \omega_j) = \frac{1}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) \right] \quad (2.48)$$

เพื่อที่จะค่าของฟังก์ชันที่ไม่รู้ $p(\mathbf{x})$ จากค่าพารามิเตอร์ $\boldsymbol{\mu}_j$ เป็นเวกเตอร์ค่าเฉลี่ยในกลุ่ม j และ Σ_j เป็น เมทริกซ์ความแปรปรวนร่วมเกี่ยว (covariance matrix) ในกลุ่ม j จากข้อมูลตัวอย่างโดยใช้เทคนิคการประมาณค่าจากข้อมูลฝึกฝน เช่น การประมาณค่ามากที่สุดที่ควรจะเป็น (maximum-likelihood) หรือ การประมาณค่าแบบเบย์ เป็นต้น ข้อดีของแนวความคิดนี้คือลักษณะเชิงความหมายง่ายต่อการเพิ่ม การลด และ การแก้ไขแบบจำลอง เมื่อจำนวนกลุ่มคำประกอบมีมากขึ้นหรือลดน้อยลง ดังรายละเอียดแสดงในบทความ [2,36,37, และ 38] แต่ข้อเสียของ

แบบจำลองสถิตินี้ ต้องใช้ข้อมูลตัวอย่างเป็นจำนวนมากสำหรับการฝึกฝนเนื่องจากแบบจำลองมีสมมุติฐานว่าลักษณะเชิงความหมายมีการแจกแจงแบบเกาส์

- **แนวคิดแบบจำลองเชิงดิสคริมิแนนต์ (Discriminant Model Approach)**

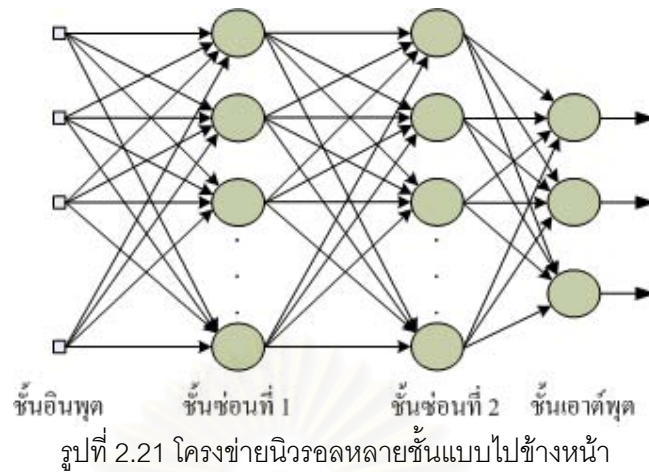
ความน่าเชื่อถือของแนวความคิดจำลองเชิงสถิติที่ขึ้นอยู่กับการฝึกฝนจำเป็นต้องมีขนาดใหญ่สำหรับแต่ละกลุ่มความหมาย แต่ในทางปฏิบัติไม่อาจจะหาลักษณะเชิงความหมายที่จำนวนมากได้ ดังนั้นการจำลองเชิงสถิติจึงไม่เหมาะสมจึงมีแนวคิดแบบจำลองเชิงดิสคริมิแนนต์นั้นคือการหาฟังก์ชันการแปลงจากเวกเตอร์ลักษณะกลาง $g_j(\mathbf{x})$ โดยฟังก์ชันอย่างง่ายเขียนอยู่ในรูปแบบของดิสคริมิแนนต์เชิงเส้นดังสมการที่ (2.49)

$$g_j(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_j + w_{j0} \quad (2.49)$$

เมื่อ \mathbf{w} เป็นค่าเวกเตอร์น้ำหนัก (weight vector) และ w_{j0} คือ ค่าไบแอส (Bias) หรือ ค่าน้ำหนักขีดเริ่ม (threshold weight) ที่คลาส j โดยที่ \mathbf{x} อยู่ในคลาส ω_j ถ้า $g_j(\mathbf{x}) > g_i(\mathbf{x})$ เมื่อ $j \neq i$ ข้อมูลฝึกฝนจะถูกนำมาสอนให้กับแบบจำลองเพื่อประมาณค่าเวกเตอร์น้ำหนักและค่าไบแอสซึ่งเทคนิคเพื่อจะหาค่าได้แก่ โคจรข่ายนิเวรอล [40] หรือ Support Vector Machine [36] เป็นต้น

2.6.1 โคจรข่ายนิเวรอล (Neural Network)

โคจรข่ายนิเวรอลมีลักษณะการคำนวณเหมือนกับการทำงานของสมองมนุษย์ ซึ่งใช้ในการวิจัยหลากหลายงานวิจัย เช่น การจดจำตัวอักษร การจดจำเสียงพูด การประมาณค่า และการจดจำแบบรูป [28] โดยที่โคจรข่ายนิเวรอลมีคุณสมบัติที่ทนต่อสัญญาณรบกวน สัญญาณผิดเพี้ยน และ สัญญาณที่ไม่สมบูรณ์ที่ได้จากในข้อมูลทางปฏิบัติ โคจรข่ายนิเวรอลหลายชั้นแบบไปข้างหน้า (feed-forward multilayer neural network) ถูกใช้งานในการสร้างตัวจำแนก มีโครงสร้างดังรูปที่ 2.21 ซึ่งเป็นการเชื่อมต่อกันหมดในแต่ละชั้นของโคจรข่าย แบ่งชั้นได้เป็นชั้นอินพุต (input layer) ชั้นซ่อน (hidden layer) และ ชั้นเอาต์พุต (output layer)



โดยค่าน้ำหนักในแต่ละการเชื่อมต่อระหว่างโหนดนิเวศจะถูกปรับโดยใช้ระเบียบวิธีการแพร่กระจายย้อนกลับ (Back-propagation algorithm) ดังสมการที่ (2.50)

$$\mathbf{w}(m+1) = \mathbf{w}(m) + \Delta\mathbf{w}(m) \quad (2.50)$$

สมการที่ (2.50) เป็นสมการที่ใช้ปรับค่าน้ำหนักในแต่ละรอบการฝึกฝน โดย $\mathbf{w}(m+1)$ เป็นค่าน้ำหนักใหม่ $\mathbf{w}(m)$ เป็นค่าน้ำหนักเดิมและ $\Delta\mathbf{w}(m)$ เป็นผลต่างค่าน้ำหนักหาได้จากสมการที่ (2.51) และ (2.52)

$$\Delta w_{ij} = \eta \delta_j o_i \quad (2.51)$$

$$\delta_j = \begin{cases} f'(net_j)(t_j - o_j) & \text{if } j \text{ is output unit} \\ f'(net_j) \sum_k \delta_k w_{jk} & \text{if } j \text{ is hidden unit} \end{cases} \quad (2.52)$$

เมื่อ $f(net_j)$ เป็น Activation function ที่หน่วย j

t_j เป็นค่าเอาต์พุตเป้าหมายที่หน่วย j

η เป็นอัตราการเรียนรู้ของโครงข่ายนิเวศ

และ o_j เป็นค่าเอาต์พุตที่ได้โครงข่ายนิเวศในรอบที่ผ่านมาที่หน่วย j

การปรับค่าน้ำหนักถูกปรับโดยใช้ข้อมูลฝึกฝนชุดหนึ่ง เมื่อค่าความผิดพลาดของค่าเอาต์พุตเป้าหมายกับค่าเอาต์พุตอยู่ในเกณฑ์ที่สามารถยอมรับได้จึงนำไปทดสอบกับข้อมูลทดสอบ จากแนวความคิดการจำแนกแบบเบย์ (bayesian classification) ถูกนำมาปรับให้ใช้กับการตีความเชิงความน่าจะเป็นของฟังก์ชันดิสคริมิแนนต์นิเวศ ซึ่งฟังก์ชันดิสคริมิแนนต์ของ

นิเวศ เมื่อ \mathbf{x} เป็นเวกเตอร์ลักษณะเฉพาะ และ w_j เป็นเมทริกซ์ค่าถ่วงน้ำหนัก (weight matrix) ที่หน่วยนิเวศ j เขียนได้เป็น

$$y_j(\mathbf{x}) = \sum_{i=1}^n w_{ij}x_i + w_{0j} \quad (2.53)$$

$$y_j(\mathbf{x}) = W_j^t \mathbf{x} + w_{0j} \quad (2.54)$$

สำหรับปัญหาการจำแนกที่มีจำนวนเป็น C กลุ่มความหมายได้จากหน่วยนิเวศจำนวน C หน่วย โดยที่ในแต่ละคลาสมีฟังก์ชันดิสคริมิแนนต์ของตัวเอง และ เวกเตอร์อินพุตใด จะกำหนดเป็นคลาส c_k ถ้า

$$y_k(\mathbf{x}) = \max_j (y_j(\mathbf{x})) \quad (2.55)$$

จากการจำแนกแบบเบย์โดยการประมาณค่าความน่าจะเป็นหลัง (Posterior Probability) มาตัดสินจากเวกเตอร์ลักษณะเฉพาะ (\mathbf{x}) ดังสมการที่ (2.56)

$$P(c_k | \mathbf{x}) = \max_j \{P(c_j | \mathbf{x})\} \quad (2.56)$$

เมื่อ $P(c_k | \mathbf{x})$ เป็นค่าความน่าจะเป็นหลังของคลาส c_k ที่ได้จาก \mathbf{x} เมื่อเปรียบเทียบสมการที่ (2.55) และ (2.56) การจำแนกแบบเบย์ในการเลือกคลาสที่เหมาะสมด้วยค่าความน่าจะเป็นหลัง เขียนอยู่ในรูปฟังก์ชันดิสคริมิแนนต์ได้เป็น

$$y_j(\mathbf{x}) = p(\mathbf{x} | c_j)P(c_j) \quad (2.57)$$

เนื่องจากขนาดของฟังก์ชันดิสคริแนนต์มีบทบาทสำคัญในการกระบวนการแก้ปัญหา ดังนั้นจึงพิจารณาฟังก์ชันดิสคริมิแนนต์ให้อยู่ในรูปลอการิทึมของความน่าจะเป็นหลัง ซึ่งจากสมการที่ (2.57) เขียนได้เป็น

$$y_j(\mathbf{x}) = \ln p(\mathbf{x} | c_j) + \ln P(c_j) \quad (2.58)$$

ซึ่งถ้าตั้งสมมุติฐานที่ว่าความหนาแน่นความน่าจะเป็นของเวกเตอร์อินพุต \mathbf{x} ในแต่ละคลาสอยู่ในรูปของการกระจายตัวแบบเกาส์ ซึ่งเขียนได้ดังสมการที่ (2.59)

$$p(\mathbf{x} | c_j) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_j)^t \Sigma_j^{-1} (\mathbf{x} - \bar{\mathbf{x}}_j)\right) \quad (2.59)$$

เมื่อ เวกเตอร์ค่าเฉลี่ย $\bar{\mathbf{x}}_j \in \mathcal{R}^n$ เป็นเวกเตอร์ขนาด n มิติ

$$\bar{\mathbf{x}}_j = E[\mathbf{x}] \quad (2.60)$$

และ $\Sigma \in \mathcal{R}^{n \times n}$ เป็นเมทริกซ์ความแปรปรวนร่วมเกี่ยว (covariance matrix) ขนาด $n \times n$

$$\Sigma_j = E\left[(\mathbf{x} - \bar{\mathbf{x}}_j)(\mathbf{x} - \bar{\mathbf{x}}_j)'\right] \quad (2.61)$$

เขียนสมการที่ (2.58) ใหม่ได้เป็น

$$y_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_j)'\Sigma_j^{-1}(\mathbf{x} - \bar{\mathbf{x}}_j) - \frac{1}{2}\ln|\Sigma_j| - \frac{n}{2}\ln 2\pi + \ln P(c_j) \quad (2.62)$$

เมื่อตัดพจน์ของค่าคงที่และให้ในทุกคลาสมีเมทริกซ์ความแปรปรวนร่วมเกี่ยวเท่ากัน $\Sigma_j = \Sigma$, $j=1, \dots, C$ ดังนั้นในพจน์ของ $\frac{1}{2}\ln|\Sigma|$ ละทิ้งได้เนื่องจากเป็นค่าคงที่ เขียนสมการใหม่ได้เป็น

$$\begin{aligned} y_j(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_j)'\Sigma^{-1}(\mathbf{x} - \bar{\mathbf{x}}_j) + \ln P(c_j) \\ &= \bar{\mathbf{x}}_j'\Sigma^{-1}\mathbf{x} - \frac{1}{2}\bar{\mathbf{x}}_j'\Sigma^{-1}\bar{\mathbf{x}}_j + \ln P(c_j) \\ &= W_j'\mathbf{x} + w_{0j} \end{aligned} \quad (2.63)$$

โดยที่สมการที่ (2.63) ละทิ้งพจน์ $\mathbf{x}'\Sigma^{-1}\mathbf{x}$ เพราะว่าในทุกเวกเตอร์ลักษณะเฉพาะเป็นอิสระออกจากกัน ดังนั้นจะได้ว่า

$$W_j' = \bar{\mathbf{x}}_j'\Sigma^{-1} \quad (2.64)$$

$$w_{0j} = -\frac{1}{2}\bar{\mathbf{x}}_j'\Sigma^{-1}\bar{\mathbf{x}}_j + \ln P(c_j) \quad (2.65)$$

ดังนั้นสรุปได้สำหรับข้อมูลที่มีการกระจายแบบเกาส์ที่มีเมทริกซ์ความแปรปรวนร่วมเกี่ยวเท่ากัน ฟังก์ชันตัดสินใจแนบของการจำแนกแบบเบย์ที่เหมาะสมจะอยู่ในรูประนาบเชิงเส้น (hyperplane) ในกรณีที่เมทริกซ์ความแปรปรวนร่วมเกี่ยวในแต่ละคลาสเป็นเมทริกซ์ไคฟังก์ชันตัดสินใจแนบของการจำแนกแบบเบย์ที่เหมาะสมอยู่ในรูประนาบกกำลังสอง (hyperquadric)

2.6.2 การวิเคราะห์ส่วนประกอบมุขสำคัญ (Principal Component Analysis)

การวิเคราะห์ส่วนประกอบมุขสำคัญ [42] ใช้ในการลดมิติที่โดยการวิเคราะห์หา ลักษณะเฉพาะร่วมกัน โดยอาศัยฟังก์ชันเกณฑ์ความผิดพลาดกำลังสอง (squared-error criterion function) ซึ่งพิจารณาได้เป็นการหาเวกเตอร์เดี่ยว (single vector \mathbf{x}_0) ที่ใช้แทนเซต ตัวอย่างขนาด n มิติ ซึ่งฟังก์ชันเกณฑ์ความผิดพลาดกำลังสองของเวกเตอร์เดี่ยว $J_0(\mathbf{x}_0)$ เขียน ดังสมการที่ (2.66)

$$J_0(\mathbf{x}_0) = \sum_{k=1}^n \|\mathbf{x}_0 - \mathbf{x}_k\|^2 \quad (2.66)$$

ค่าเวกเตอร์เดี่ยวที่ดีที่สุดจะทำให้ J_0 มีค่าน้อยที่สุด ในกรณีอย่างง่ายเมื่อสมมติให้เวกเตอร์เดี่ยว เป็นค่าเวกเตอร์เฉลี่ย $\mathbf{x}_0 = \mathbf{m}$ เมื่อ \mathbf{m} เป็นเวกเตอร์ค่าเฉลี่ยดังสมการที่ (2.67)

$$\mathbf{m} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad (2.67)$$

จะได้ว่า

$$\begin{aligned} J(\mathbf{x}_0) &= \sum_{k=1}^n \|(\mathbf{x}_0 - \mathbf{m}) - (\mathbf{x}_k - \mathbf{m})\|^2 \\ &= \sum_{k=1}^n \|\mathbf{x}_0 - \mathbf{m}\|^2 - 2 \sum_{k=1}^n (\mathbf{x}_0 - \mathbf{m})' (\mathbf{x}_k - \mathbf{m}) + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \\ &= \sum_{k=1}^n \|\mathbf{x}_0 - \mathbf{m}\|^2 - 2(\mathbf{x}_0 - \mathbf{m})' \sum_{k=1}^n (\mathbf{x}_k - \mathbf{m}) + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \\ &= \sum_{k=1}^n \|\mathbf{x}_0 - \mathbf{m}\|^2 + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \end{aligned} \quad (2.68)$$

จากสมการที่ (2.68) เนื่องจากผลรวมในพจน์ที่ 2 เป็นอิสระจาก \mathbf{x}_0 ซึ่งเป็นการ เป็นทำให้น้อยที่สุด (minimized) ด้วยการเลือก $\mathbf{x}_0 = \mathbf{m}$ นั่นคือการหาเวกเตอร์ \mathbf{m} ที่ทำให้ค่า น้อยที่สุด

เนื่องจากเวกเตอร์ค่าเฉลี่ยเป็นเวกเตอร์ตัวแทนมีจำนวนมิติเป็นศูนย์มิติทำให้ไม่ เหมาะกับข้อมูลที่มีการเปลี่ยนแปลงได้ ดังนั้นถ้าสร้างเส้นตรงเส้นหนึ่งทีผ่านเวกเตอร์ค่าเฉลี่ยถือว่า เป็นการฉายเวกเตอร์ (vector projection) ให้กับเวกเตอร์เดี่ยวที่มีปริภูมิเป็น 1 มิติ โดยให้ \mathbf{e} เป็น เวกเตอร์ 1 หน่วยในทิศทางไปตามเส้นตรงแล้วสมการของเส้นตรงเขียนได้เป็น

$$\mathbf{x} = \mathbf{m} + a\mathbf{e} \quad (2.69)$$

เมื่อค่าสเกลาร์ a เป็นค่าจำนวนจริง ซึ่งเป็นระยะทาง (distance) ระหว่าง เวกเตอร์ \mathbf{x} กับ เวกเตอร์ค่าเฉลี่ย \mathbf{m} ดังนั้นสามารถเซตสัมประสิทธิ์ a_k ที่เหมาะสมได้จากฟังก์ชันเกณฑ์ความผิดพลาดที่น้อยที่สุดจากสมการที่ (2.66)

$$\begin{aligned} J_1(a_1, \dots, a_k, \mathbf{e}) &= \sum_{k=1}^n \|(\mathbf{m} + a_k \mathbf{e}) - \mathbf{x}_k\|^2 = \sum_{k=1}^n \|a_k \mathbf{e} - (\mathbf{x}_k - \mathbf{m})\|^2 \\ &= \sum_{k=1}^n a_k^2 \|\mathbf{e}\|^2 - 2 \sum_{k=1}^n a_k \mathbf{e}^T (\mathbf{x}_k - \mathbf{m}) + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \end{aligned} \quad (2.70)$$

จาก $\|\mathbf{e}\| = 1$ และเมื่อผลต่างอนุพันธ์ย่อยของสมการที่ (2.70) กำหนดให้มีค่าเป็น 0 ซึ่งเป็นการหาเซตสัมประสิทธิ์ a_k ที่ทำให้ฟังก์ชันเกณฑ์ความผิดพลาดกำลังมีค่าน้อยที่สุดจะได้

$$a_k = \mathbf{e}^T (\mathbf{x}_k - \mathbf{m}) \quad (2.71)$$

จากสมการที่ (2.71) อธิบายเชิงเรขาคณิตได้ว่าผลที่ได้เป็นผลเฉลยกำลังสองที่น้อยสุดที่ด้วยการเวกเตอร์ \mathbf{x}_k ไปบนทิศทาง \mathbf{e} ซึ่งผ่านเวกเตอร์ค่าเฉลี่ย \mathbf{m}

ปัญหาลำดับถัดมาคือการหาทิศทางเวกเตอร์ \mathbf{e} ที่ดีที่สุด ซึ่งแก้ปัญหานี้ได้โดย จากเมทริกซ์การกระจาย (scatter Matrix) \mathbf{S} ที่นิยามเป็น

$$\mathbf{S} = \sum_{k=1}^n (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T \quad (2.72)$$

เมื่อแทนสมการที่ (2.71) ลงในสมการที่ (2.70) จะได้

$$\begin{aligned} J_1(\mathbf{e}) &= \sum_{k=1}^n a_k^2 - 2 \sum_{k=1}^n a_k^2 + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \\ &= - \sum_{k=1}^n [\mathbf{e}^T (\mathbf{x}_k - \mathbf{m})]^2 + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \\ &= - \sum_{k=1}^n \mathbf{e}^T (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T \mathbf{e} + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \\ &= -\mathbf{e}^T \mathbf{S} \mathbf{e} + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \end{aligned} \quad (2.73)$$

เวกเตอร์ \mathbf{e} ที่ทำให้ J_1 น้อยที่สุดก็ต่อเมื่อ $\mathbf{e}^T \mathbf{S} \mathbf{e}$ เป็นค่าที่มากที่สุด โดยให้วิธีผลคูณลากรางจ์ (Lagrange multipliers) ให้ λ เป็น ตัวคูณเทียบ (undetermined multiplier) ฟังก์ชันลากรางจ์เขียนได้เป็น

$$u = \mathbf{e}^T \mathbf{S} \mathbf{e} - \lambda \mathbf{e}^T \mathbf{e} \quad (2.74)$$

เมื่อหาอนุพันธ์ของสมการที่ (2.74) ด้วย \mathbf{e} จะได้

$$\frac{\partial u}{\partial \mathbf{e}} = 2\mathbf{S}\mathbf{e} - 2\lambda\mathbf{e} \quad (2.75)$$

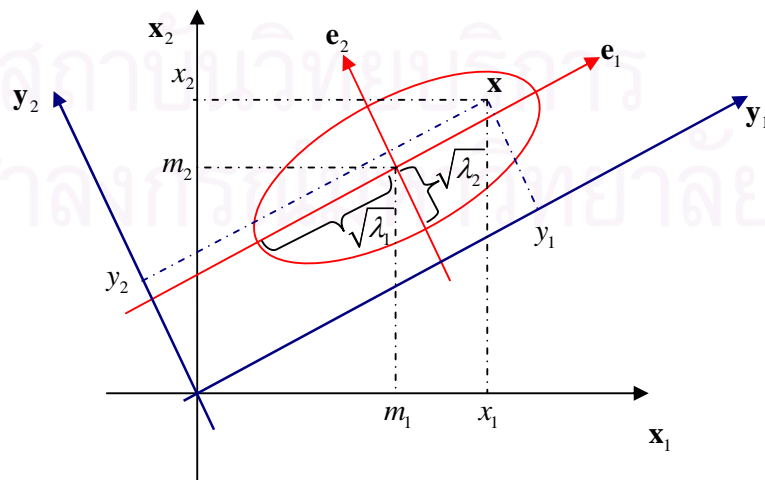
เพื่อกำหนดให้เวกเตอร์เกรเดียนต์เป็นศูนย์จะเห็นได้ว่าเวกเตอร์ \mathbf{e} เขียนอยู่ในรูปแบบการวิเคราะห์เวกเตอร์ลักษณะเฉพาะ (eigenvector-eigenvalue analysis) ของเมทริกซ์การกระจายเป็น

$$\mathbf{S}\mathbf{e} = \lambda\mathbf{e} \quad (2.76)$$

เพราะว่า $\mathbf{e}^T \mathbf{S}\mathbf{e} = \lambda \mathbf{e}^T \mathbf{e} = \lambda$ ดังนั้นเราควรเลือก เวกเตอร์ทิศทาง \mathbf{e} หรือเวกเตอร์ลักษณะเฉพาะได้จากค่าลักษณะเฉพาะ (eigenvalue) ที่มากที่สุด กล่าวอีกนัยหนึ่งได้ว่าการฉายข้อมูล (projection) ลงบนเวกเตอร์มิติใหม่ใน 1 มิติ ข้อมูลจะถูกฉายลงไปบนเส้นที่ผ่านเวกเตอร์ค่าเฉลี่ยในทิศทางของเวกเตอร์ลักษณะที่ให้ค่าลักษณะเฉพาะมากที่สุดของเมทริกซ์การกระจาย ดังนั้นสามารถขยายผลที่ได้จากการฉายเวกเตอร์ 1 มิติเป็นการฉายเวกเตอร์ d' มิติจากสมการที่ (2.69) เขียนได้เป็น

$$\mathbf{x} = \mathbf{m} + \sum_{i=1}^{d'} a_i \mathbf{e}_i \quad (2.77)$$

เมื่อ $d' \leq d$ โดยที่ d เป็นจำนวนมิติ \mathbf{x}_k และ a_i เป็นส่วนประกอบของ \mathbf{x} ของในแต่ละฐานสำคัญ \mathbf{e}_i และเรียกว่า a_i ส่วนประกอบमुखสำคัญ ถ้าวิเคราะห์ในเชิงเรขาคณิต ถ้าข้อมูลจำนวน d อยู่ในรูปของกลุ่มของวงรีหลายระนาบ (hyperellipsoidal) แล้ว ค่าเวกเตอร์ลักษณะเฉพาะจะอยู่บนแต่ละแกนของวงรีในแต่ละระนาบดังรูปที่ 2.22



รูปที่ 2.22 การแปลงข้อมูลตัวอย่างลงบนส่วนประกอบमुखสำคัญใน 2 มิติ

ดังนั้นจากรูปที่ 2.22 เขียนอยู่ในรูปของการแปลงเมทริกซ์ได้ดังสมการที่ (2.78)

$$\mathbf{y} = \mathbf{e}^T \mathbf{x} \quad (2.78)$$

เมื่อหาค่าเวกเตอร์เฉลี่ยและเมทริกซ์ความแปรปรวนร่วม¹ ของ \mathbf{y} ได้ดังสมการที่ (2.79) และ (2.80) ตามลำดับ

$$\mathbf{m}_y = E\{\mathbf{e}^T \mathbf{x}\} = \mathbf{e}^T \mathbf{m} \quad (2.79)$$

$$S_y = E\{(\mathbf{y} - \mathbf{m}_y)(\mathbf{y} - \mathbf{m}_y)^T\} = \mathbf{e}^T S \mathbf{e} \quad (2.80)$$

จากรูปแบบของการวิเคราะห์เวกเตอร์ลักษณะเฉพาะในสมการที่ (2.74) และ สามารถจัดสมการได้ใหม่เป็น

$$\mathbf{e}^T S \mathbf{e} = \lambda I = \Lambda \quad (2.81)$$

ดังนั้นจากสมการที่ (2.80) และสมการที่ (2.81) เมทริกซ์ความแปรปรวนร่วมของ \mathbf{y} เขียนอยู่ในรูปของเมทริกซ์ทแยงมุม $\Lambda = \lambda I$ ดังนั้นสามารถอธิบายได้ว่าการวิเคราะห์ส่วนประกอบสำคัญเป็นการแปลงเชิงเส้นที่ทำให้สร้างเมทริกซ์ทแยงมุมในพิกัดใหม่ นั้นหมายความว่าในแต่ละเวกเตอร์ในแต่แกนของพิกัดใหม่ไม่มีสหสัมพันธ์กัน (uncorrelation) และเป็นอิสระออกจากกัน

2.6.3 การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นบนปริภูมิศูนย์ (Null Space based Linear Discriminant Analysis)

ตัวจำแนกโดยทั่วไปจะอาศัยหลักพื้นฐานการแทนคุณสมบัติลักษณะเวกเตอร์เฉพาะที่ดึงได้จากข้อมูลอินพุต เช่น ข้อมูลตัวอักษร ข้อมูลภาพ หรือ ข้อมูลเสียง เป็นต้น กล่าวคือ การหาเวกเตอร์ลักษณะเฉพาะที่มีความสำคัญในแต่ละกลุ่มที่ต้องการจำแนกมาเป็นตัวแทนข้อมูลในแต่ละกลุ่ม อย่างไรก็ตามเมื่อเราใช้ข้อมูลอินพุตในแต่ละกลุ่มอย่างจำกัด หรือ ข้อมูลอินพุตที่ได้มานั้นไม่ก่อให้เกิดประโยชน์ต่อการจำแนก เราจำเป็นต้องเลือกเวกเตอร์เฉพาะย่อยที่มีนัยสำคัญจากเวกเตอร์ลักษณะเฉพาะอินพุตหรือเป็นการลดจำนวนลดขนาดของจำนวนเวกเตอร์ลง เพื่อให้ผลการจำแนกดีขึ้น ซึ่งเทคนิคการลดขนาดของเวกเตอร์ลงนี้เรียกว่า การฉายปริภูมิย่อย (subspace projection)

¹ หาได้เหมือนกับเมทริกซ์การกระจายดังสมการที่ (2.72)

เทคนิคการฉายปริภูมิย่อยที่เป็นที่รู้จักกันมี 2 เทคนิค คือ การวิเคราะห์ส่วนประกอบมุมค่าสำคัญที่ได้กล่าวไว้ในหัวข้อมาผ่านมา และ การวิเคราะห์ดิสคริมิแนนต์เชิงเส้น (linear discriminant analysis – LDA) สำหรับการวิเคราะห์ส่วนประกอบมุมค่าสำคัญ นั้นจะเลือกปริภูมิย่อยที่ข้อมูลมีการกระจายใกล้เคียงกับปริภูมิเดิมมากที่สุด นั่นคือให้ค่าเวกเตอร์ลักษณะเฉพาะมากกว่าศูนย์ซึ่งเรียกว่าแกนส่วนประกอบมุมค่าสำคัญ (principal component axis) โดยที่เวกเตอร์ที่ให้ค่าลักษณะเฉพาะเข้าใกล้ศูนย์เรียกว่าแกนศูนย์ (null axis) ซึ่งปริภูมิย่อยที่เกิดจากแกนศูนย์จะถูกเรียกว่าปริภูมิศูนย์ (null space) ในขณะที่การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นจะเลือกปริภูมิย่อยที่เหมาะสมกับการจำแนกมากที่สุด นั่นคือการกระจายภายในกลุ่ม (within-class scatter) น้อย และการกระจายระหว่างกลุ่ม (between-class scatter) ที่มาก โดยที่ทั้งสองเทคนิคนี้ประสบความสำเร็จในงานการรู้จำใบหน้า [43-45]

แนวความคิดพื้นฐานของการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นเป็นการหาการแปลงเชิงเส้นเพื่อที่จะได้กลุ่มของลักษณะที่สามารถแยกแยะได้หลังจากการแปลง ซึ่งเริ่มต้นจากการนิยามและตั้งสมมุติฐานทางคณิตศาสตร์ ให้ n แทนด้วยจำนวนมิติของปริภูมิของข้อมูลตัวอย่าง ตั้งต้น และ c เป็นจำนวนกลุ่มที่ต้องการจำแนก เมทริกซ์การกระจายระหว่างกลุ่ม และ เมทริกซ์การกระจายภายในกลุ่ม นิยามได้เป็น

$$S_b = \sum_{i=1}^c N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T = \Phi_b \Phi_b^T \quad (2.82)$$

$$S_w = \sum_{i=1}^c \sum_{k \in C_i} (\mathbf{x}_k - \mathbf{m}_i)(\mathbf{x}_k - \mathbf{m}_i)^T = \Phi_w \Phi_w^T \quad (2.83)$$

เมื่อ N_j เป็นจำนวนของตัวอย่างในกลุ่มที่ C_j ($j=1,2,\dots,c$)

N เป็นจำนวนของตัวอย่างทั้งหมด

\mathbf{m}_j เป็นเวกเตอร์ค่าเฉลี่ยภายในกลุ่มที่ C_j

\mathbf{m} เป็นเวกเตอร์ค่าเฉลี่ยทั้งหมดของตัวอย่างทั้งหมด

Φ_b เป็นเมทริกซ์ผลต่างเวกเตอร์ค่าเฉลี่ยภายในกลุ่ม \mathbf{m}_j กับ เวกเตอร์ค่าเฉลี่ยทั้งหมดของตัวอย่างข้อมูล \mathbf{m}

และ Φ_w เป็นเมทริกซ์ผลต่างเวกเตอร์ภายในกลุ่ม \mathbf{x}_j กับ เวกเตอร์ค่าเฉลี่ยภายในกลุ่มที่ C_j

โดยเมทริกซ์การกระจายทั้งหมด หรือ เมทริกซ์ความแปรปรวนร่วมทั้งหมดนิยามได้เป็น

$$S_i = S_b + S_w = \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T = \Phi_i \Phi_i^T \quad (2.84)$$

การวิเคราะห์หาค่าลักษณะเฉพาะเชิงเส้นเป็นการหาการฉายที่เหมาะสมที่สุด $W = [w_1 w_2 \dots w_{c-1}]$ จากเงื่อนไขที่ว่า

$$J(W) = \arg \max_w \frac{|W^T S_b W|}{|W^T S_w W|} \quad (2.85)$$

เรียกกฎเกณฑ์นี้ว่า “Fisher criterion function” แล้ว w หาได้จากการหาผลเฉลยจาก ปัญหาค่าสุดท้ายฟังก์ชันการแยกแยะเชิงเส้น ของการวิเคราะห์หาค่าลักษณะเฉพาะเชิงเส้นเขียนได้เป็น

$$d_i(\mathbf{x}) = W^T (\mathbf{x} - \mathbf{m}_i) \quad (2.86)$$

2.6.3.1 การคำนวณการฉายที่เหมาะสม

ผลเฉลยของการแปลงที่เหมาะสมของสมการที่ (2.85) อธิบายได้ว่าการแปลงที่เหมาะสมต้องทำให้เมทริกซ์การกระจายระหว่างกลุ่ม (S_b) ถูกแปลงให้มีการกระจายที่มากขึ้น ในขณะที่เดียวกันต้องทำให้เมทริกซ์การกระจายภายในกลุ่ม (S_w) ถูกแปลงให้มีการกระจายที่น้อยลง ดังนั้นในการแก้ปัญหาการแปลงที่เหมาะสมทำได้ 2 วิธีคือ

- การคำนวณการวิเคราะห์เวกเตอร์ลักษณะเฉพาะของเมทริกซ์ $S_w^{-1} S_b$ [46] แต่เทคนิคนี้ผลที่ได้มีค่าตัวเลขที่ไม่แน่นอนเนื่องจากขึ้นอยู่กับค่าของเมทริกซ์ผกผันของเมทริกซ์การกระจายระหว่างภายในกลุ่ม S_w ที่เป็นเมทริกซ์ที่มีขนาดใหญ่ โดยส่วนใหญ่ในทางปฏิบัติการหาการแปลงที่เหมาะสมจะหาได้จากการแปลงที่ทำให้เมทริกซ์การกระจายระหว่างกลุ่ม และ เมทริกซ์การกระจายภายในกลุ่ม เป็นเมทริกซ์ที่ทแยงมุมพร้อมกัน (simultaneous diagonalization)

- การคำนวณการแปลงโดยทำให้เมทริกซ์เป็นเมทริกซ์ที่ทแยงมุมพร้อมกัน [47] ซึ่งเราหาการแปลงเชิงเส้นที่สามารถทำให้ S_b และ S_w เป็นเมทริกซ์ที่ทแยงมุมพร้อมกันได้ จากกระบวนการดังต่อไปนี้

เริ่มจากการหาการแปลงแบบไวเทนนิ่ง (whitening transformation) ของ S_w คือการหาเมทริกซ์การแปลงที่ทำให้ S_w เป็นเมทริกซ์เอกลักษณ์ (identity matrix) ดังสมการที่ (2.87)

$$\mathbf{y} = \Theta^{-1/2} \Phi \mathbf{x} \quad (2.87)$$

เมื่อ Θ และ Φ เป็นค่าเวกเตอร์ลักษณะเฉพาะและเวกเตอร์ลักษณะเฉพาะของ S_w ตามลำดับ โดยที่

$$S_w \Phi = \Phi \Theta \quad \text{และ} \quad \Phi^T \Phi = I \quad (2.88)$$

แล้ว S_w และ S_b ถูกแปลงไปเป็น

$$\Theta^{-1/2} \Phi^T S_w \Phi \Theta^{-1/2} = I \quad (2.89)$$

$$\Theta^{-1/2} \Phi^T S_b \Phi \Theta^{-1/2} = K \quad (2.90)$$

โดยทั่วไป K ไม่เป็นเมทริกซ์ทแยงมุม

ขั้นที่ 2 หากการแปลงเชิงตั้งฉากปกติ (orthonormal transformation) เพื่อให้ K เป็นเมทริกซ์ทแยงมุม นั่นคือ

$$\mathbf{z} = \Psi^T \mathbf{y} \quad (2.91)$$

เมื่อ Ψ และ Λ เป็นค่าเวกเตอร์ลักษณะเฉพาะและเวกเตอร์ลักษณะเฉพาะของ K แล้ว

$$K \Psi = \Psi \Lambda \quad \text{และ} \quad \Psi^T \Psi = I \quad (2.92)$$

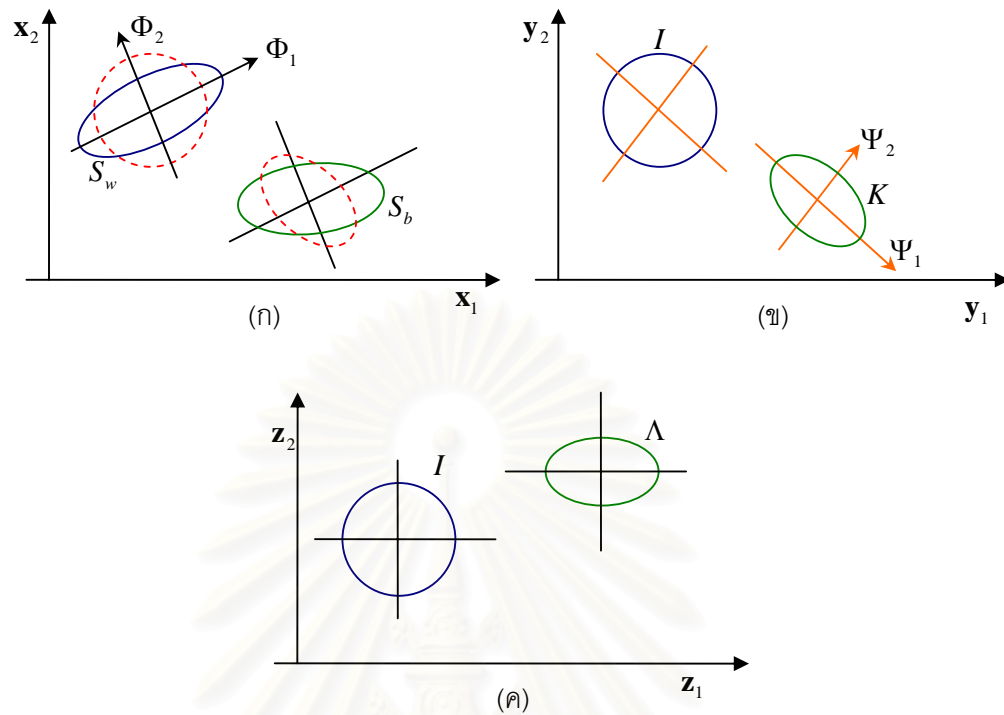
จากสมการที่แปลงเมทริกซ์ที่ (2.89) จะไม่เปลี่ยนแปลงเมื่อถูกแปลงดังนี้

$$\text{จากสมการที่ (2.89) เป็น} \quad \Psi^T I \Psi = \Psi^T \Psi = I \quad (2.93)$$

$$\text{จากสมการที่ (2.90) เป็น} \quad \Psi^T K \Psi = \Lambda \quad (2.94)$$

ดังนั้นเมทริกซ์การกระจายทั้งสองจะเป็นเมทริกซ์ทแยงมุม ดังรูปที่ 2.23 แสดงตัวอย่างการแปลงเมทริกซ์ปริภูมิ 2 มิติ โดยที่การแปลงรวมเขียนได้เป็น

$$W^* = \Phi \Theta^{1/2} \Psi \quad (2.95)$$



รูปที่ 2.23 การทำให้เมทริกซ์ทแยงมุมพร้อมกัน

กระบวนการข้างต้นเป็นการหาการแปลงที่ทำให้เมทริกซ์ทแยงมุมพร้อมกัน โดยเริ่มจากการหาการแปลงที่ทำให้เมทริกซ์ S_w เป็นเมทริกซ์เอกลักษณะเพื่อหลีกเลี่ยงการหาเมทริกซ์ผกผันไม่ได้ของ S_w ในทางกลับกันเราสามารถหาการแปลงที่ทำให้เมทริกซ์ S_b เป็นเมทริกซ์เอกลักษณะก่อน เพื่อรักษาข้อมูลที่เป็นประโยชน์บนเมทริกซ์การกระจาย S_b เมื่อละทิ้งปริภูมิสูญศูนย์ของ S_b ซึ่งการทำให้เมทริกซ์ทแยงมุมพร้อมกันโดยเริ่มจาก S_b เรียกว่าการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบตรง (direct LDA) [46]

เงื่อนไขในการเลือกปริภูมิหลัก (principal space) ที่เหมาะสมจากเงื่อนไข $\frac{\lambda_i}{\lambda_{\max}} > \epsilon$ [45] ถ้าเงื่อนไขไม่เป็นจริงแล้วจะเรียกว่าปริภูมิสูญศูนย์ (null space) ซึ่งเราจะเรียกปริภูมิหลักจากเวกเตอร์ลักษณะเฉพาะที่ได้จากการแปลงที่เกิดจาก S_b และเลือกปริภูมิสูญศูนย์จากการแปลงที่เกิดจาก S_w เมทริกซ์การแปลงที่ได้เขียนดังสมการที่ (2.92)

ในบทนี้ได้กล่าวถึงทฤษฎีและความรู้พื้นฐานต่างๆ เกี่ยวกับโครงสร้างวิดิทัศน์ซึ่งในการค้นคืนวิดิทัศน์หรือการให้คำประกอบวิดิทัศน์จำเป็นต้องมีการแยกส่วนวิดิทัศน์ทางเวลาซึ่งจะได้ข้อดีวิดิทัศน์ ข้อดีวิดิทัศน์คือว่าเป็นหน่วยที่เล็กที่สุดของลำดับวิดิทัศน์ จากนั้นจึงดึงลักษณะเฉพาะระดับล่างจากข้อดีวิดิทัศน์ แบ่งได้ 2 ลักษณะเฉพาะ ได้แก่ ลักษณะเฉพาะเชิงพื้นที่

ซึ่งดึงลักษณะเฉพาะจากเฟรมสำคัญ และ ลักษณะเฉพาะเชิงเวลาได้จากเวกเตอร์ลักษณะการเคลื่อนที่ของกล้อง เป็นต้น โดยที่ลักษณะเฉพาะได้มาจากกรรมวิธีประมวลผลภาพเหล่านี้ถือว่าเป็นลักษณะเฉพาะระดับล่างของวิดีโอ เพื่อนำมาสร้างแบบจำลองเชิงความหมายในการจำแนกประเภทกีฬาโดยใช้เทคนิคการรู้จำแบบรู้ ส่วนเทคนิคที่นำเสนอในวิทยานิพนธ์นี้จะกล่าวต่อไปในบทที่ 3 และ ผลการทดลองเทคนิคที่นำเสนอในบทที่ 4



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

เทคนิคที่นำเสนอ

ในการจำแนกวีดิทัศน์ประเภทกีฬาให้มีประสิทธิภาพจะประกอบด้วยส่วนสำคัญ 2 ส่วน คือ การเลือกเวกเตอร์ลักษณะเฉพาะที่เหมาะสม ซึ่งใช้เป็นตัวแทนของวีดิทัศน์แต่ละประเภทกีฬา ตัวอย่างเช่น ลักษณะเฉพาะสี รูปร่าง และ การเคลื่อนที่ เป็นต้น และการเลือกใช้ตัวจำแนกที่เหมาะสม โดยที่ตัวจำแนกนั้นเป็นแบบจำลองเพื่อเชื่อมต่อระหว่างเวกเตอร์ลักษณะเฉพาะระดับสูง หรือ ลักษณะเชิงความหมาย (ในที่นี้หมายถึง “ประเภทกีฬา”) กับเวกเตอร์ลักษณะระดับล่าง

การเลือกเวกเตอร์ลักษณะเฉพาะที่เหมาะสมเพื่อแยกแยะประเภทกีฬานั้น ลักษณะอย่างหนึ่งที่ได้เห็นได้ชัดเจน คือ ลักษณะเฉพาะสี เนื่องจากกีฬาในแต่ละประเภทมีความแตกต่างกันของสีสนามในการแข่งขัน ลักษณะเฉพาะขอบ ซึ่งเป็นโครงร่างของวัตถุในวีดิทัศน์ที่สามารถมาแยกแยะความแตกต่างได้ และ ลักษณะเฉพาะการเคลื่อนที่ของกล้อง ซึ่งในแต่ละชนิดกีฬามีการเคลื่อนที่ของกล้องที่แตกต่างกัน ส่วนลักษณะพื้นผิวนั้นเนื่องจากลักษณะสนามกีฬาโดยส่วนใหญ่ประกอบด้วยบริเวณที่เป็นเนื้อเดียวกัน (homogeneous regions) ดังนั้นเป็นเวกเตอร์ลักษณะเฉพาะที่ไม่เหมาะสมกับการจำแนก

ส่วนการเลือกใช้ตัวจำแนกที่เหมาะสมแล้วนั้นจะทำให้ผลของการจำแนกได้ความถูกต้องและมีประสิทธิภาพที่ตรงกับประเภทกีฬาที่กำหนดไว้ โดยเฉพาะอย่างยิ่งเมื่อจำนวนของประเภทกีฬาเพิ่มมากขึ้นการเลือกใช้ตัวจำแนกที่เหมาะสมมีบทบาทสำคัญอย่างยิ่ง เนื่องจากความคล้ายคลึงของประเภทกีฬาจะมีมากยิ่งขึ้นด้วย

ดังนั้นวิทยานิพนธ์นี้จึงได้มีแนวความคิด 2 ประการ คือ ประการแรกเป็นการวิเคราะห์หาเวกเตอร์ลักษณะเฉพาะที่เหมาะสมกับประเภทกีฬา โดยลักษณะเฉพาะที่นำมาวิเคราะห์ได้แก่ ลักษณะเฉพาะสี เช่น ฮิสโตแกรมสี และ ฮอโตคออร์ริโลแกรม ลักษณะการเคลื่อนที่กล้อง และ ลักษณะขอบสี เพื่อค้นหาลักษณะเฉพาะที่มีความสามารถในการจำแนกประเภทวีดิทัศน์กีฬาได้ ซึ่งจะกล่าวหัวข้อที่ 3.1 ประการที่สอง การนำลักษณะเฉพาะที่มีความสามารถจำแนกประเภทวีดิทัศน์กีฬาได้มาปรับปรุงความถูกต้องของการจำแนกโดยใช้ตัวจำแนกโครงข่ายนิรเวอร์ลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ในหัวข้อที่ 3.2 และ ปรับปรุงความถูกต้องของการ

จำแนกโดยใช้ปริภูมิผู้ศูนย์ ในหัวข้อที่ 3.3 โดยข้อมูลวีดิทัศน์กีฬาที่ใช้ในวิทยานิพนธ์ฉบับนี้บันทึกจากการแพร่สัญญาณโทรทัศน์โดยทั่วไปในประเทศไทย ซึ่งบันทึกอยู่ในฐานข้อมูลวีดิทัศน์ในรูปแบบของ MPEG-1 ที่ความละเอียดของภาพเป็น 384x448 จุดภาพ อัตรากรอบภาพ (frame rate) เป็น 25 เฟรมต่อวินาที

3.1 การวิเคราะห์ความถูกต้องของการจำแนกประเภทวีดิทัศน์กีฬาโดยใช้ลักษณะระดับล่างของวีดิทัศน์กีฬา

ลักษณะเฉพาะที่เหมาะสมกับการจำแนกประเภทวีดิทัศน์กีฬาที่ได้นำมาใช้ในวิทยานิพนธ์นี้ประกอบด้วย

3.1.1 ลักษณะเฉพาะสี

ในวิทยานิพนธ์นี้เลือกใช้แบบจำลองสี RGB ในการประมวลผลภาพสำหรับการดึงลักษณะเฉพาะสี เนื่องจากในการแสดงผลภาพโดยทั่วไปใช้แบบจำลองสี RGB โดยจำนวนสีภายในภาพหลังจากควอนไทซ์ที่ใช้ คือ 64 ถึงสี ลดลงมาจาก 224 สี หรือ ประมาณ 16.7 ล้านสี สำหรับการแสดงผลภาพที่ขนาด 24 บิตต่อจุดภาพ ซึ่งค่านี้นิยมใช้ในงานวิจัยเป็นส่วนใหญ่ [27] เนื่องจากการแยกแยะความแตกต่างของระดับค่าสีของสายตามนุษย์นั้นมีความละเอียดไม่มากนัก และ วิธีในการดึงลักษณะเฉพาะของสีที่เลือกใช้ ได้แก่ ฮิสโตแกรมสี และ ออโตคอรัลโลแกรม

- ฮิสโตแกรมสี

สำหรับการวิเคราะห์ฮิสโตแกรมสีในวิทยานิพนธ์นี้โดยแบ่งกลุ่มของสีออกเป็น 64 สี เพื่อใช้แทนลักษณะสีในแต่ละประเภทกีฬา ซึ่งเป็นความน่าจะเป็นของสีที่พบภายในภาพจากสมการที่ (2.13) โดยที่ เวกเตอร์ของฮิสโตแกรมสีเขียนได้ดังสมการที่ (3.1)

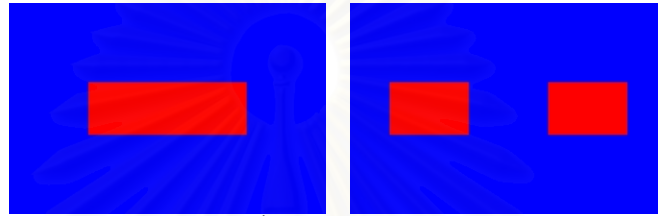
$$H(I) = (h_{c_1}, h_{c_2}, \dots, h_{c_m}) \quad (3.1)$$

โดยที่ m เป็นจำนวนสีภายในภาพหลังการควอนไทซ์สี

- ออโตคอรัลโลแกรม

เพื่อปรับปรุงความถูกต้องของการจำแนกประเภทวีดิทัศน์กีฬา สำหรับวิธีออโตคอรัลโลแกรมสามารถทำได้โดยการเลือกค่าพารามิเตอร์ในการคำนวณค่าออโตคอรัลโลแกรมของภาพในวีดิทัศน์ที่เหมาะสม ซึ่งได้แก่ จำนวนสีภายในภาพหลังจากการควอนไทซ์สี และ เซตของระยะห่างระหว่างจุดภาพ 2 จุดภาพใดๆ โดยค่าพารามิเตอร์ทั้งสองนี้จะส่งผลต่อขนาดของ

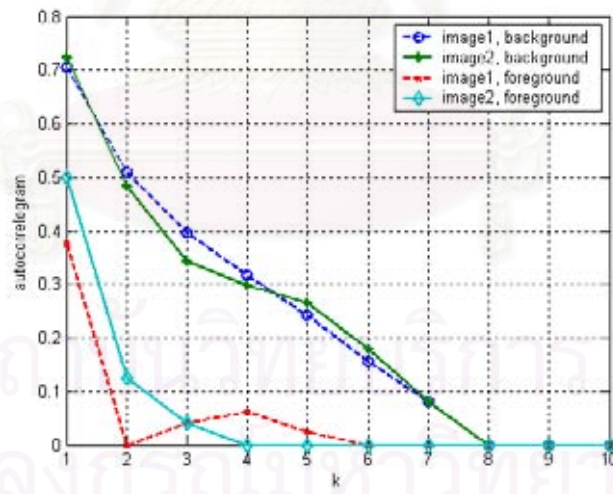
เวกเตอร์ลักษณะเฉพาะ และ ความซับซ้อนรวมทั้งระยะเวลาที่ใช้ในการคำนวณออโตคอรรีโลแกรมของภาพในการสร้างเวกเตอร์ลักษณะเฉพาะ และ สิ่งสำคัญอีกประการหนึ่งคือขนาดของภาพที่นำมาใช้ในการทดลองเนื่องจากขนาดของเวกเตอร์ลักษณะเฉพาะที่ใช้แทนภาพทั้งภาพ เมื่อภาพมีขนาดใหญ่จำเป็นต้องใช้เวกเตอร์ลักษณะเฉพาะที่มีขนาดใหญ่ตามขนาดของภาพ ซึ่งแสดงได้ดังตัวอย่างดังรูปที่ 3.1 ในกรณีแรกเป็นภาพที่มีขนาดเล็กขนาด 8×8 จุดภาพ และ ในกรณีที่ 2 เป็นภาพที่มีขนาดใหญ่ 200×100 จุดภาพ



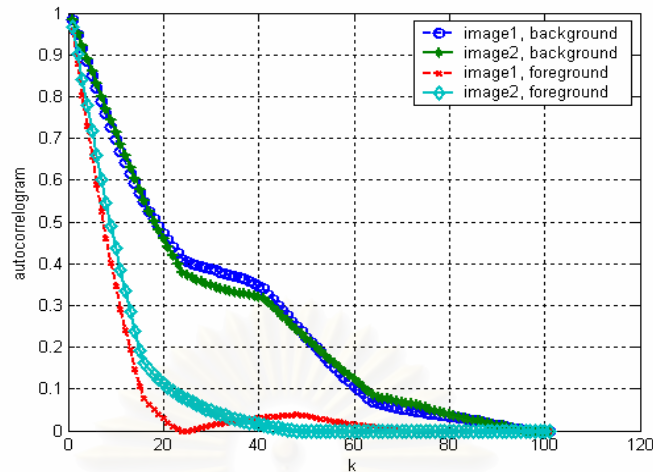
รูปที่ 3.1 ภาพตัวอย่าง

(ก) ภาพที่ 1 (image 1) (ข) ภาพที่ 2 (image 2)

ผลของออโตคอรรีโลแกรมของทั้งสองกรณีแสดงได้ดังรูปที่ 3.2 และ รูปที่ 3.3 ตามลำดับ



รูปที่ 3.2 ออโตคอรรีโลแกรมของภาพขนาด 8×8 จุดภาพ



รูปที่ 3.3 ออโตคอรีโกลแกรมของภาพขนาด 200x100 จุดภาพ

จากรูปที่ 3.2 และ รูปที่ 3.3 เห็นได้ว่าภาพที่มีขนาดแตกต่างกันขนาดของเวกเตอร์ลักษณะออโตคอรีโกลแกรมที่ใช้ไม่เท่ากัน ในขณะที่ภาพที่ใช้มีขนาดใหญ่ มีความจำเป็นต้องใช้เซตของระยะห่างระหว่างจุดภาพที่มีขนาดใหญ่เพื่อเป็นตัวแทนภาพได้ทั้งภาพ เมื่อใช้เซตของระยะห่างจุดภาพที่มีขนาดใหญ่แล้วทำให้ขนาดของเวกเตอร์ลักษณะเฉพาะมีขนาดใหญ่ตามมาด้วย ดังนั้นเพื่อต้องการจำแนกประเภทวัตถุที่สนใจให้ถูกต้องแล้วขนาดเวกเตอร์ลักษณะเฉพาะที่ต้องการจำเป็นต้องมีขนาดเล็ก และ ความสะดวกและรวดเร็วในการประมวลผลของการสร้างตัวจำแนกต่อไป จึงลดขนาดของภาพลง 448x384 จุดภาพ เป็น 112x96 จุดภาพ และ กำหนดเซตของระยะห่างระหว่างจุดภาพ 2 จุดภาพใดๆ เท่ากับ $k = \{1, 3, 5, 7, 9\}$ ซึ่งคำนวณออโตคอรีโกลแกรมได้จากสมการที่ (2.17) และ เขียนเวกเตอร์ลักษณะเฉพาะแทนออโตคอรีโกลแกรมได้ดังสมการที่ (3.2)

$$\gamma_{C_i, C_i}^{(k)}(l) = \begin{bmatrix} \gamma_{c_1, c_1}^{(1)} & \gamma_{c_1, c_1}^{(2)} & \cdots & \gamma_{c_1, c_1}^{(k)} \\ \gamma_{c_2, c_2}^{(1)} & \gamma_{c_2, c_2}^{(2)} & \cdots & \gamma_{c_2, c_2}^{(k)} \\ \gamma_{c_m, c_m}^{(1)} & \gamma_{c_m, c_m}^{(2)} & \cdots & \gamma_{c_m, c_m}^{(k)} \end{bmatrix} \quad (3.2)$$

ขนาดของเวกเตอร์แทนออโตคอรีโกลแกรมจะเท่ากับ $m \times d$

โดยที่ m เป็นจำนวนสีภายหลังการควอนไทซ์สี

k คือ ระยะห่างระหว่างจุดภาพ 2 จุดภาพใดๆ

และ d เป็นจำนวนสมาชิกของเซตระยะห่างระหว่าง 2 จุดภาพใดๆ

และจัดเวกเตอร์เป็น 1 มิติได้เป็น

$$\gamma_{C_i, C_i}^{(k)}(I) = [\gamma_{c_1, c_1}^{(1)} \gamma_{c_2, c_2}^{(1)} \dots \gamma_{c_m, c_m}^{(1)} \gamma_{c_1, c_1}^{(2)} \gamma_{c_2, c_2}^{(2)} \dots \gamma_{c_m, c_m}^{(2)} \gamma_{c_1, c_1}^{(k)} \gamma_{c_2, c_2}^{(k)} \dots \gamma_{c_m, c_m}^{(k)}] \quad (3.3)$$

ส่วนความซับซ้อนและเวลาที่ใช้ในการคำนวณออกโตคอร์รีโโลแกรมของภาพจะแปรตามขนาดของภาพและเซตของระยะห่างระหว่าง 2 จุดภาพที่ใช้ ซึ่งเมื่อลดขนาดของภาพแล้วทำให้ความซับซ้อนและเวลาในการคำนวณน้อยลง

3.1.2 ลักษณะเฉพาะการเคลื่อนที่ของกล้อง

การวิเคราะห์ลักษณะการเคลื่อนที่ของกล้อง เนื่องจากกีฬาแตกต่างกัน ลักษณะการเคลื่อนที่ของกล้องมีความแตกต่างกัน โดยวิทยานิพนธ์นี้ได้กำหนดลักษณะการเคลื่อนที่ของกล้องให้แต่ละเฟรมของวิดีโอที่ได้จากสมการที่ (2.27) เป็น 8 รูปแบบได้แก่ กล้องไม่มีการเคลื่อนที่ การเคลื่อนที่ของกล้องสายไปทางขวา (0 องศา) การเคลื่อนที่ของกล้องสายไปทางขวา (180 องศา) การเคลื่อนที่ของกล้องก้ม (180 องศา) การเคลื่อนที่ของกล้องเงย (90 องศา) การเคลื่อนที่ของกล้องในแนว 45 องศา การเคลื่อนที่ของกล้องในแนว 135 องศา การเคลื่อนที่ของกล้องในแนว 225 องศา และ การเคลื่อนที่ของกล้องในแนว 315 องศา นอกจากนี้ยังใช้ขนาดเวกเตอร์การเคลื่อนที่ในแต่ละเฟรมที่ได้จากสมการที่ (2.24) เป็นเวกเตอร์ลักษณะของการเคลื่อนที่ โดยในกำหนดให้ทุกข้อต่อจะมีขนาด 100 เฟรม โดยกำหนดลักษณะการเคลื่อนที่ ทุก 5 เฟรม ซึ่งจะได้เวกเตอร์ลักษณะเฉพาะขนาด 40 มิติ ซึ่งเขียนเวกเตอร์ลักษณะเฉพาะของการเคลื่อนที่ได้ดังสมการที่ (3.4)

$$Motion = [S^1 V^1 S^2 V^2 \dots S^n V^n] \quad S^n \in \{No_motion, 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 315^\circ\} \quad (3.4)$$

เมื่อ n เป็นจำนวนเฟรมภายในข้อต่อในที่นี้เป็นการหาลักษณะการเคลื่อนที่ทุก 5 เฟรม และ V^n เป็นขนาดเวกเตอร์การเคลื่อนที่ที่เฟรม n

3.1.3 ลักษณะเฉพาะขอบสี

กีฬาแตกต่างกัน รูปร่างหรือลักษณะของขอบจะมีความแตกต่างกัน โดยวิทยานิพนธ์นี้เลือกใช้ตัวพรางแบบ sobel เนื่องจากใช้ขนาดตัวพรางที่มีขนาดใหญ่จึงแสดงผลกระทบจากสัญญาณรบกวนได้ดี และ สร้างเวกเตอร์ลักษณะขอบสีโดยตรวจหาขอบในแต่ละองค์ประกอบของสีในปริภูมิสี YUV เพื่อแยกแยะเฟรมของวิดีโอที่มีลักษณะขอบเหมือนกันใน

องค์ประกอบความส่องสว่าง นอกจากนี้เวกเตอร์ลักษณะของขอบสีเป็นลักษณะที่มีความหมาย โดยแบ่งภาพออกเป็นบล็อก แล้วกำหนดความหมายของขอบในแต่ละบล็อกซึ่งความหมายของขอบแบ่งได้เป็น 5 ความหมาย คือ ขอบในแนวนอน (-) ขอบในแนวตั้ง (|) ขอบในแนว 45 องศา (/) และ ขอบในแนว 135 องศา (\) และไม่มีขอบ ในแต่ละบล็อกถูกกำหนดความหมายขอบโดยใช้ฮิสโตแกรมของขอบในแต่ละบล็อกที่ให้จำนวนความหมายมากที่สุด ซึ่งขนาดของเวกเตอร์ลักษณะของขอบสีจะขึ้นอยู่กับจำนวนบล็อกที่กำหนดให้กับภาพ ดังนั้นเราจึงลดขนาดของภาพจาก 448x384 จุดภาพ เป็น 112x96 จุดภาพ เพื่อความสะดวกและรวดเร็วในการประมวลผลของการสร้างตัวจำแนก ซึ่งขนาดบล็อกที่ใช้เป็น 16x16 จุดภาพ ขนาดของเวกเตอร์ลักษณะเฉพาะขอบสีที่ได้เป็น 126 (7x6x3) มิติ จากสมการที่ (2.44) และ (2.45) เขียนเป็นเวกเตอร์ลักษณะเฉพาะได้เป็น

$$E(I) = (E^{B_1} E^{B_2} \dots E^{B_N}) \quad E^{B_k} \in \{-, |, /, \backslash, null\} \quad (3.5)$$

$$edgecolor(I) = [E^Y \ E^U \ E^V] \quad (3.6)$$

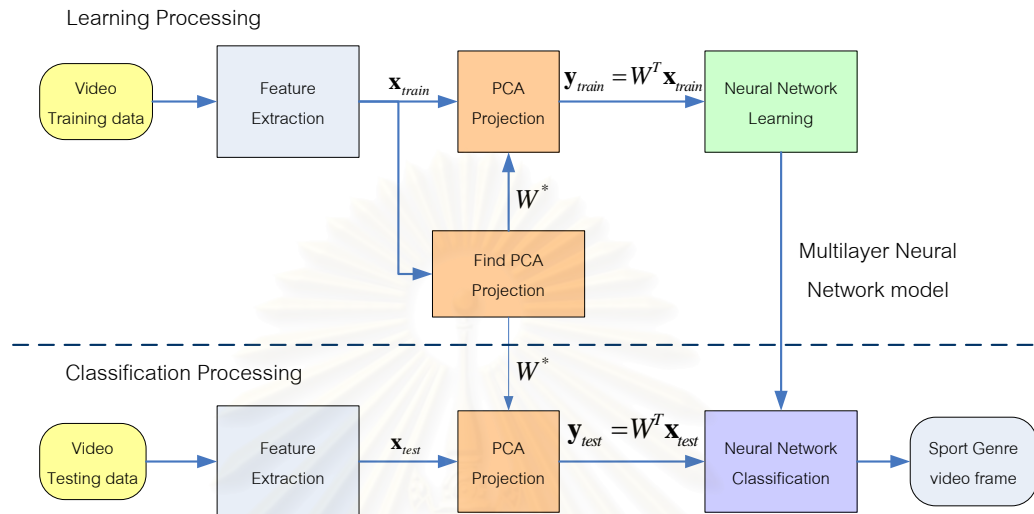
เมื่อ N เป็นจำนวนบล็อกที่แบ่งให้กับเฟรมภายในข้อต่อ

3.2 การปรับปรุงความแม่นยำของการจำแนกประเภทวิดีโอทัศนียภาพโดยใช้ตัวจำแนกโครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ในวิทยานิพนธ์นี้นำเสนอการจำแนกโดยใช้โครงข่ายนิวรอล เนื่องจากโครงข่ายนิวรอลเป็นเทคนิคในการสร้างแบบจำลองหรือการสร้างตัวจำแนกที่ง่ายและได้ผลการจำแนกที่ดีในหลากหลายการประยุกต์การใช้งาน เพียงแค่ออกแบบจำนวนหน่วยนิวรอลอินพุตของโครงข่ายนิวรอลให้เท่ากับจำนวนมิติของเวกเตอร์ลักษณะเฉพาะ ส่วนจำนวนหน่วยนิวรอลเอาต์พุตของโครงข่ายนิวรอลจะเท่ากับจำนวนประเภทกีฬาที่ต้องการจำแนก โครงข่ายนิวรอลสามารถถูกออกแบบให้มีความสามารถในการจำแนกข้อมูลที่มีความซับซ้อนโดยการเพิ่มจำนวนหน่วยนิวรอลในชั้นซ่อนของโครงข่ายนิวรอล นั้นหมายความว่าถ้าข้อมูลตัวอย่างสามารถแบ่งได้ด้วยฟังก์ชันเชิงเส้น ฉะนั้นจะเลือกใช้โครงข่ายนิวรอลที่ปราศจากชั้นซ่อน แต่ถ้าข้อมูลตัวอย่างไม่สามารถแบ่งได้ด้วยฟังก์ชันเชิงเส้น เราเพียงเพิ่มชั้นซ่อนให้กับโครงข่ายนิวรอลและหาจำนวนหน่วยนิวรอลที่เพียงพอสำหรับจำแนกข้อมูลตัวอย่างได้

เนื่องจากข้อมูลตัวอย่างที่มีความซับซ้อนประกอบไปด้วยข้อมูลที่มีประโยชน์และไม่มีประโยชน์ต่อการจำแนกประเภทวิดีโอทัศนียภาพ ดังนั้นเพื่อลดความซับซ้อนในกระบวนการฝึกฝนโครงข่ายนิวรอลและดึงเอาเฉพาะข้อมูลที่มีนัยสำคัญ อีกทั้งช่วยลดขนาดปริภูมิข้อมูลตัวอย่างเพื่อให้มีประสิทธิภาพต่อการจำแนก ดังนั้นจึงใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ

ร่วมกับโครงข่ายนิวรอล ซึ่งถือว่าการวิเคราะห์ส่วนประกอบमुखสำคัญเป็นกรรมวิธีประมวลผลเริ่มต้น (pre-processing) ให้กับข้อมูลตัวอย่าง ขั้นตอนการจำแนกประเภทวีดิทัศน์กีฬาโดยใช้โครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ แสดงได้ดังรูปที่ 3.4



รูปที่ 3.4 แผนภาพขั้นตอนการจำแนกประเภทวีดิทัศน์โดยใช้โครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

3.2.1 การวิเคราะห์ส่วนประกอบमुखสำคัญ

จากรูปที่ 3.4 แสดงแผนภาพขั้นตอนของการนำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิวรอล ซึ่งขั้นตอนแรกคือขั้นตอนดึงลักษณะเฉพาะจากวีดิทัศน์ ขั้นตอนต่อมาคือการหาเมทริกซ์การแปลงที่เหมาะสม W^* สำหรับการแปลงเวกเตอร์เวกเตอร์ลักษณะเฉพาะ ลงบนปริภูมิส่วนประกอบमुखสำคัญที่ได้จากข้อมูลฝึกฝนโดยทั้งข้อมูลฝึกฝน (training data) และ ข้อมูลทดสอบ (testing data) จะถูกแปลงลงบนปริภูมิส่วนประกอบमुखสำคัญด้วยเมทริกซ์การแปลงเดียวกันดังสมการที่ (3.6)

$$\mathbf{y} = \mathbf{W}^{*T} (\mathbf{x} - \boldsymbol{\mu}) \quad (3.7)$$

โดยที่ $\boldsymbol{\mu}$ เป็นเวกเตอร์ค่าเฉลี่ยของข้อมูลฝึกฝน และ เวกเตอร์ \mathbf{y} เป็นเวกเตอร์ที่จะนำไปใช้กับโครงข่ายนิวรอลต่อไป ซึ่งการหาเมทริกซ์การแปลง \mathbf{W}^{*T} ที่เหมาะสมหาได้จากกระบวนการดังต่อไปนี้

- วิเคราะห์เวกเตอร์ลักษณะเฉพาะ (eigen-analysis) จากเมทริกซ์ความแปรปรวนร่วมเกี่ยวของข้อมูลฝึกฝน Σ_x ซึ่งจะได้ค่าลักษณะเฉพาะ $(\lambda_1, \lambda_2, \dots, \lambda_n)$ ที่ตรงกับเวกเตอร์

ลักษณะเฉพาะ $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ ในที่นี้กำหนดให้ค่าลักษณะเฉพาะเรียงจากมากไปหาน้อยเป็น $\lambda_1 > \lambda_2 > \dots > \lambda_n$

- ถ้ากำหนดให้เมทริกซ์การแปลงจาก n มิติไปสู่ปริภูมิส่วนประกอบमुखสำคัญ d มิติ เมื่อ $d \leq n$ แล้วเมทริกซ์การแปลงที่เหมาะสม เขียนได้อยู่ในการเรียงกันในแนวหลักของเวกเตอร์ลักษณะเฉพาะแรก ดังสมการที่ (3.7)

$$W^* = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_d]_{n \times d} \quad (3.8)$$

3.2.2 ตัวจำแนกโครงข่ายนิเวรอล

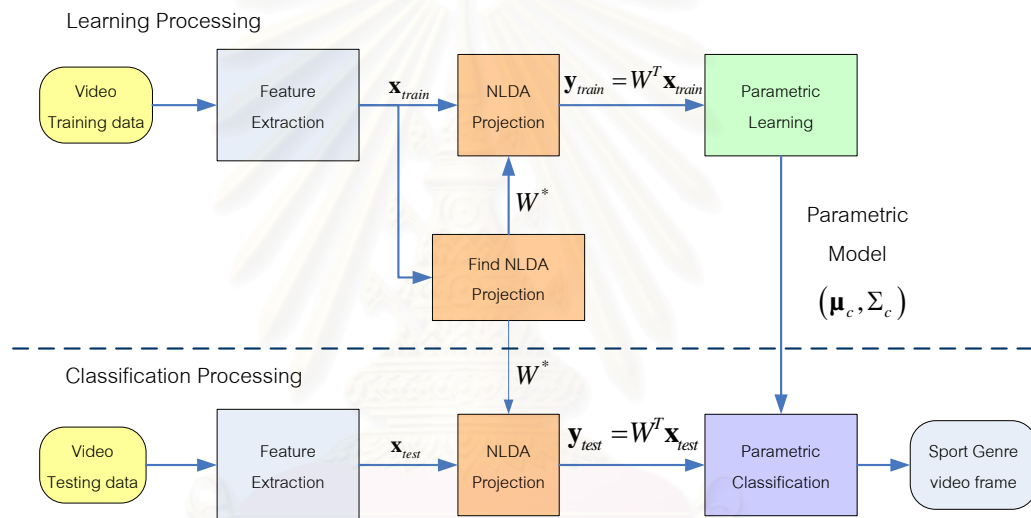
หลังจากการแปลงข้อมูลฝึกฝนลงปริภูมิส่วนประกอบमुखสำคัญแล้ว เราออกแบบโครงสร้างให้กับโครงข่ายนิเวรอลเป็นแบบหนึ่งโครงสร้างสำหรับหลายประเภทกีฬา โดยกำหนดให้จำนวนหน่วยนิเวรอลอินพุตเป็น d หน่วยนิเวรอลซึ่งเท่ากับการขนาดของปริภูมิส่วนประกอบमुखสำคัญ และจำนวนนิเวรอลเอาต์พุตเป็น c หน่วยนิเวรอลเท่ากับจำนวนชนิดกีฬาที่ต้องการจำแนก นอกจากนี้กำหนดให้ฟังก์ชันการกระตุ้น (activation function) เป็น tangent hyperbolic transfer function สำหรับหน่วยนิเวรอลที่อยู่ในชั้นซ่อน และ กำหนดให้ ฟังก์ชันการกระตุ้นสำหรับหน่วยนิเวรอลเอาต์พุตเป็น log sigmoid transfer function ซึ่งเป็น softmax function เพื่อประมาณค่าความน่าจะเป็น โดยที่เอาต์พุตจากหน่วยนิเวรอลในชั้นสุดท้ายนี้แทนความน่าจะเป็นในแต่ละประเภทกีฬา

เมื่อออกแบบโครงสร้างของโครงข่ายนิเวรอลแล้วในวิทยานิพนธ์นี้กำหนดให้เทคนิคการปรับค่าเวกเตอร์ถ่วงน้ำหนักของนิเวรอลเป็นแบบการเรียนรู้แบบปรับตัวได้ (adaptive gradient descent learning) เนื่องจากให้ผลการเรียนรู้ที่รวดเร็วและให้ผลที่ดีเท่ากับการปรับค่าเวกเตอร์ถ่วงน้ำหนักแบบ gradient descent โดยทั่วไป ด้วยโดยกำหนดให้อัตราการเรียนรู้ 0.001 และ เพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น

เพื่อปรับปรุงความถูกต้องของการจำแนกให้มีความสามารถทางด้าน Generalization ดังนั้นกระบวนการฝึกฝนจึงแบ่งตัวอย่างข้อมูลฝึกฝนออกเป็น 2 ข้อมูล คือ ข้อมูลเรียนรู้ (learning data) และ ข้อมูลวาลิเดต (validation data) ด้วยอัตราส่วน 4 ต่อ 1 โดยข้อมูลวาลิเดต ใช้สำหรับการตัดสินใจหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลตรวจสอบความสมเหตุสมผลมีแนวโน้มเพิ่มขึ้น ถึงแม้ว่าความผิดพลาดบนข้อมูลเรียนรู้จะลดลงก็ตาม ทั้งนี้เพื่อหลีกเลี่ยงปรากฏการณ์ “ท่องจำ” หรือ overfitting บนโครงข่ายนิเวรอล กฎการตัดสินใจของตัวจำแนกนิเวรอลนี้เป็นไปตามประเภทของกีฬาที่ให้ความน่าจะเป็นมากที่สุด

3.3 การปรับปรุงความถูกต้องของการจำแนกประเภทวิดีโอทัศนกีฬาโดยใช้ตัวจำแนกปริภูมิสูงศูนย์

เพื่อปรับปรุงของความถูกต้องของการจำแนกให้มีประสิทธิภาพ วิทยานิพนธ์นี้ได้นำเสนอเทคนิคอีกเทคนิคหนึ่งซึ่งเรียกว่า ตัวจำแนกโดยใช้ปริภูมิสูงศูนย์ เนื่องจากปริภูมิสูงศูนย์เป็นปริภูมิที่เกิดจากเมทริกซ์การแปลงที่ทำให้การกระจายตัวระหว่างกลุ่มมีระยะห่างกันมากที่สุด ในขณะที่การกระจายตัวภายในกลุ่มชนิดเดียวกันเข้าใกล้กันมากที่สุด นอกจากนี้ปริภูมิสูงศูนย์มาจากการประยุกต์ fisher criteria กับข้อมูลที่จำนวนน้อย เมื่อเทียบกับขนาดของปริภูมิ นั้นหมายความว่าสามารถจำแนกข้อมูลได้ง่ายขึ้นเมื่อสร้างตัวจำแนกในปริภูมิสูงศูนย์ ซึ่งขั้นตอนการจำแนกโดยใช้ปริภูมิสูงศูนย์แสดงได้ดังรูปที่ 3.5



รูปที่ 3.5 แผนภาพขั้นตอนการจำแนกประเภทวิดีโอทัศนกีฬาโดยใช้ตัวจำแนกปริภูมิสูงศูนย์

ขั้นตอนแรกเหมือนกับในหัวข้อที่ 3.2 คือการดึงลักษณะเฉพาะของวิดีโอ แต่แตกต่างกันในขั้นตอนของการหาเมทริกซ์การแปลงที่เหมาะสม W^* ที่ใช้กฎเกณฑ์ "Fisher criterion function" ในหัวข้อที่ 2.6.3 เมื่อหาเมทริกซ์การแปลงที่เหมาะสมข้อมูลฝึกฝนและข้อมูลทดสอบจะถูกแปลงบนปริภูมิสูงศูนย์ ซึ่งในการหาเมทริกซ์การแปลงที่เหมาะสมมีขั้นตอนในการหาดังต่อไปนี้

- หาส่วนประกอบमुखสำคัญจากเวกเตอร์ค่าเฉลี่ยและเมตริกซ์ความแปรปรวนทั้งหมดจาก ซึ่งจะได้เมทริกซ์การแปลงบนปริภูมิส่วนประกอบमुखสำคัญเป็น $(W^{PCA})_{n \times L}$ โดยเลือกขนาดของปริภูมิส่วนประกอบमुखสำคัญ (L) จากเงื่อนไข $\frac{\lambda_i}{\lambda_{max}} > \epsilon$

- ฉายข้อมูลฝึกฝนทั้งหมดบนปริภูมิส่วนประกอบमुखสำคัญโดยใช้เมตริกซ์การแปลงจากขั้นตอนที่ผ่านมา แล้วหาเวกเตอร์ค่าเฉลี่ยในแต่ละกลุ่มและเมตริกซ์ความแปรปรวนร่วมเกี่ยวในแต่ละกลุ่ม
- หาเมตริกซ์การแปลงบนปริภูมิผู้ศูนย์จากการวิเคราะห์เวกเตอร์ลักษณะเฉพาะจากเมตริกซ์ความแปรปรวนร่วมเกี่ยวในแต่ละกลุ่ม S_w แล้วจะได้เมตริกซ์การแปลงบนปริภูมิผู้ศูนย์ $(N_w)_{L \times M}$ จากเงื่อนไข $\frac{\lambda_i}{\lambda_{\max}} < \epsilon$ เมื่อ M คือขนาดปริภูมิผู้ศูนย์
- $W^* = W^{PCA} N_w$ เป็นเมตริกซ์การแปลงที่เหมาะสม

ทั้งนี้สามารถนำปริภูมิผู้ศูนย์ร่วมกับโครงข่ายนิเวรอลก็ได้ แต่ที่ไม่ใช้เนื่องจากการฉายลงปริภูมิผู้ศูนย์ กลุ่มของข้อมูลที่อยู่ในกลุ่มเดียวกันจะกองกันอยู่ที่จุดเดียว ดังนั้นจึงใช้ตัวจำแนกที่สร้างจากแบบจำลองทางสถิติเพียงพอ จึงไม่จำเป็นต้องใช้โครงข่ายนิเวรอล ซึ่งตัวจำแนกที่สร้างขึ้นเป็นตัวจำแนกแบบเชิงสถิติโดยการหาระยะห่างระหว่างตัวแทนกลุ่มที่เป็นไปตามสมการที่ (3.8)

$$f(x) = \arg \min_{k=1, \dots, c} \left\| (m'_k - W^{*T} x)^T \Sigma'_k (m'_k - W^{*T} x) \right\| \quad (3.9)$$

เมื่อข้อมูลทดสอบถูกฉายบนปริภูมิผู้ศูนย์เป็น $x' = W^{*T} x$ โดยที่เวกเตอร์ค่าเฉลี่ยในแต่ละกลุ่มประเภทกีฬาเป็น $m'_k = W^{*T} m_k$ และ Σ'_k เป็นเมตริกซ์ฝึกฝนในแต่ละกลุ่ม k ของเมตริกซ์ความแปรปรวนร่วมเกี่ยว $\Sigma_k = \sum_{x' \in c_k} (x' - m'_k)(x' - m'_k)^T$

ในบทนี้ได้กล่าวเทคนิคที่นำเสนอในวิทยานิพนธ์ฉบับนี้ประกอบด้วยการศึกษาหาคุณลักษณะเฉพาะที่เหมาะสมกับการจำแนกประเภทกีฬา และ นำเสนอการนำการวิเคราะห์ส่วนประกอบमुखสำคัญมาใช้ร่วมกับโครงข่ายนิเวรอล นอกจากนี้ยังนำเสนอการใช้ปริภูมิผู้ศูนย์เพื่อปรับปรุงการจำแนกประเภทวิถีทัศน์กีฬา ซึ่งจะกล่าวในผลการทดลองในบทต่อไป

จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการทดลอง

เนื้อหาในบทนี้จะกล่าวถึง การทดสอบการจำแนกประเภทวิถีทัศนิกภาพด้วย
เวกเตอร์ลักษณะเฉพาะระดับล่างต่างๆ ได้แก่ เวกเตอร์ลักษณะเฉพาะฮิสโตแกรมสี เวกเตอร์
ลักษณะเฉพาะฮิสโตแกรมสีเวกเตอร์ลักษณะเฉพาะการเคลื่อนที่ของกล้อง และ เวกเตอร์
ลักษณะเฉพาะขอบสี โดยวัดประสิทธิภาพของการจำแนกด้วยร้อยละความแม่นยำของการจำแนก
ประเภทวิถีทัศนิกภาพ และ คอนฟิวชันเมทริกซ์ (confusion matrix) โดยรายละเอียดของการทดลอง
แบ่งได้เป็น 2 ส่วนใหญ่คือ ส่วนที่หนึ่ง ประกอบด้วยการทดลองที่ 1 ถึง การทดลองที่ 4 เป็นการ
ตรวจสอบเวกเตอร์คุณลักษณะระดับล่างแบบใดที่เหมาะสมกับการจำแนกประเภทวิถีทัศนิกภาพ
และส่วนการทดลองที่ 5 เป็นการวิเคราะห์ตัวจำแนกที่ได้ออกแบบไว้จากการทดลองที่ 1 ถึง 4 เมื่อ
จำนวนประเภทกึ่งภาพมีจำนวนเพิ่มมากขึ้น และ การทดลองที่ 6 เป็นการทดลองเพื่อปรับปรุง
ประสิทธิภาพของการจำแนกเมื่อจำนวนประเภทวิถีทัศนิกภาพมีมากขึ้นโดยการเลือกใช้ตัวจำแนก
ในปริภูมิสูงสู่น้อยในการจำแนก ซึ่งมีรายละเอียดทั้งหมดดังต่อไปนี้

4.1 การทดลองที่ 1 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิถีทัศนิกภาพของ ลักษณะเฉพาะฮิสโตแกรมสี โดยใช้โครงข่ายนิรอรลร่วมกับการวิเคราะห์ส่วนประกอบमुख สำคัญ

การทดลองแรกเป็นการศึกษาและวิเคราะห์เกี่ยวกับการจำแนกประเภทวิถีทัศนิก
ภาพด้วยลักษณะเฉพาะฮิสโตแกรมสีสำหรับการจำแนกประเภทวิถีทัศนิกภาพ อีกทั้งเพื่อทดสอบ
การวิเคราะห์ส่วนประกอบमुखสำคัญมาใช้ร่วมกับโครงข่ายนิรอรลแบบต่าง ได้แก่ แบบไม่มีชั้นซ่อน
และ แบบมีชั้นซ่อน สามารถทำให้ร้อยละความแม่นยำเพิ่มขึ้นได้หรือไม่

วัตถุประสงค์

เพื่อทดสอบประสิทธิภาพความแม่นยำของการจำแนกประเภทวิถีทัศนิกภาพโดยใช้
โครงข่ายนิรอรลแบบต่างๆ ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะฮิส
โตแกรมสี

ปัจจัยกำหนดในการทดลอง

- โครงสร้างของโครงข่ายนิวรอลซึ่งออกแบบไว้ในการทดลองแบ่งได้เป็น 4 กรณีได้แก่
 - กรณีที่ 1 แบบไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 2 แบบไม่มีชั้นซ่อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 3 แบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 4 แบบมี 1 ชั้นซ่อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
- อัตราการเรียนรู้ เป็น 0.0001 และ เพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น สำหรับทั้ง 4 โครงข่ายนิวรอลที่ได้ออกแบบไว้
- ประเภทกีฬาสำหรับการทดลองเป็น 4 ประเภทกีฬา ได้แก่ มวยไทย เทนนิส ฟุตบอล และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปตวันักกีฬา และ ฉากผู้ชม รวมเป็นจำนวน 14 วิดิทัศน์ คิดเป็นเวลาประมาณ 12 ชั่วโมง ซึ่งดึงเฟรมสำคัญทุกๆ หนึ่งเฟรมจาก 100 เฟรม จะได้ภาพในแต่ละประเภทกีฬาเป็น 1500 ภาพ รวมทั้งสิ้น 6000 ภาพ ขนาดของภาพเป็น .384x448 จุดภาพ ซึ่งขนาดเวกเตอร์ลักษณะเฉพาะฮิสโตแกรมสีมีขนาด เป็น 64 มิติ
- อัตราส่วนระหว่างจำนวนข้อมูลฝึกฝนต่อจำนวนข้อมูลทดสอบเป็น 2 ต่อ 1
- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาเลดิเต (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาเลดิเต (Validate data) มีแนวโน้มสูงขึ้น
- ทดลองในแต่ละโครงข่ายนิวรอลจำนวน 5 ครั้งแล้วจึงนำมาหาค่าเฉลี่ยของร้อยละความแม่นยำและคอนฟิวชันเมทริกซ์เฉลี่ย

ผลการทดลอง

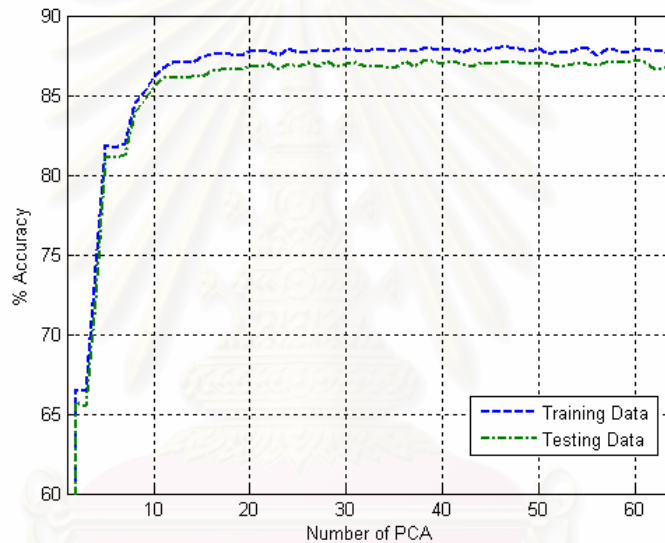
ผลการทดลองแบ่งออกเป็น 4 กรณีดังนี้

1. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ ซึ่งแสดงได้ตารางที่ 4.1

ตารางที่ 4.1 ความแม่นยำของการจำแนกด้วยโครงข่ายแบบไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะฮิสโตแกรมสี

	ข้อมูลฝึกฝน	ข้อมูลทดสอบ
ร้อยละความแม่นยำ	87.31	86.56

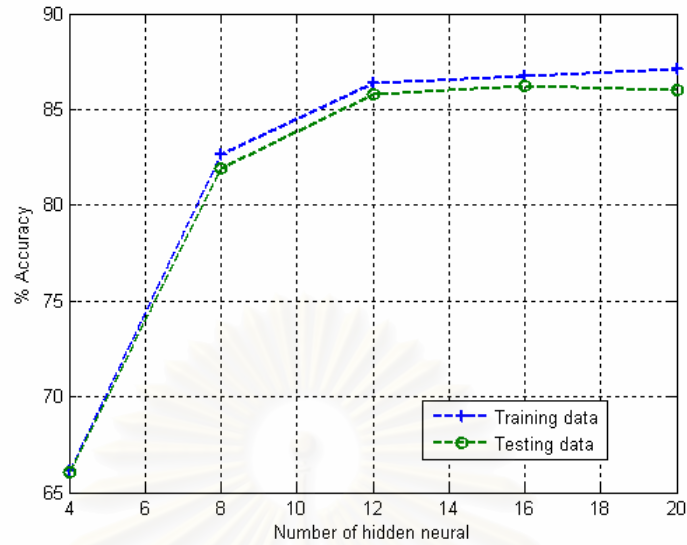
2. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญซึ่งได้ผลการทดลองดังรูปที่ 4.1



รูปที่ 4.1 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับลักษณะเฉพาะฮิสโตแกรมสี

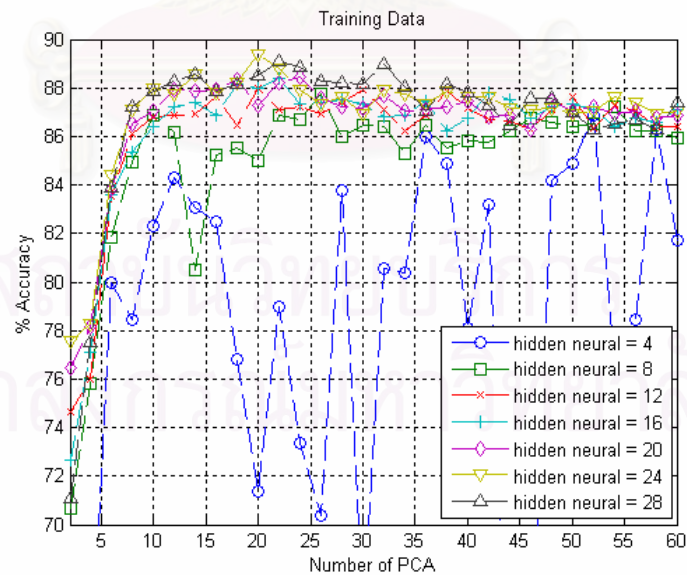
3. ผลการทดลองโครงข่ายที่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม โดยการเพิ่มจำนวนหน่วยนิเวศในชั้นซ่อนซึ่งได้ผลการทดลองดังรูปที่ 4.2

จุฬาลงกรณ์มหาวิทยาลัย

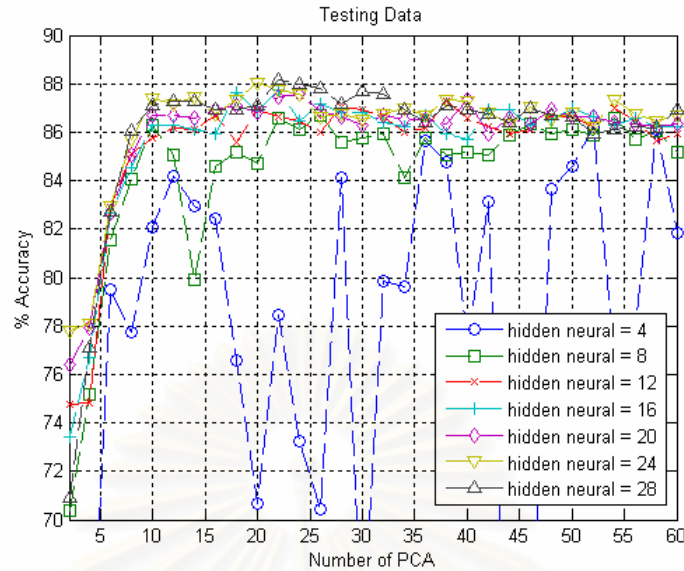


รูปที่ 4.2 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะฮิสโตแกรมสี

4. ผลการทดลองโครงข่ายนิวรอลชั้นซ่อน 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดปริภูมิส่วนประกอบमुखสำคัญ และ เพิ่มจำนวนหน่วย นิวรอลในชั้นซ่อน ซึ่งผลการทดลองได้ดังรูปที่ 4.3 และ รูปที่ 4.4 สำหรับชุดฝึกฝนและชุดทดสอบ ตามลำดับ



รูปที่ 4.3 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน



รูปที่ 4.4 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลอง

จากการทดลองกับโครงสร้างโครงข่ายนิรวลทั้ง 4 โครงสร้าง จากลักษณะเฉพาะฮิสโตแกรมสี สำหรับกรณีที่ 1 ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญรวม ผลความแม่นยำของข้อมูลทดสอบดังตารางที่ 4.1 สำหรับ ลักษณะเฉพาะฮิสโตแกรมสี ประมาณร้อยละ 86.56 เมื่อเปรียบเทียบกับกรณีที่ 2 นำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิรวลที่ไม่มีชั้นซ่อนดังรูปที่ 4.1 พบว่าเมื่อจำนวนปริภูมิของการวิเคราะห์ส่วนประกอบमुखสำคัญที่นำมาใช้มีจำนวนมากขึ้น ความแม่นยำของความถูกต้องของการจำแนกมีแนวโน้มมากขึ้น จนกระทั่งจำนวนปริภูมิของการวิเคราะห์ส่วนประกอบमुखสำคัญมากกว่า 20 มิติ แนวโน้มร้อยละของความแม่นยำมีแนวโน้มคงที่ประมาณร้อยละ 87.10 ซึ่งได้ผลที่ดีกว่าในกรณีที่ 1 เมื่อทดลองในกรณีที่ 3 ที่ไม่ได้้นำการวิเคราะห์ส่วนประกอบमुखสำคัญเข้าร่วมแต่เพิ่มชั้นซ่อน 1 ชั้น จากนั้นทดลองเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อน พบว่าเมื่อจำนวนหน่วยนิรวลในระดับชั้นซ่อนมีจำนวนเพิ่มมากขึ้นร้อยละความแม่นยำมีแนวโน้มเพิ่มมากขึ้น จากนั้นจะคงที่เมื่อจำนวนหน่วยนิรวลมากกว่า 16 หน่วย ดังรูปที่ 4.2 ซึ่งได้ร้อยละความแม่นยำประมาณร้อยละ 86.10 และกรณีที่ 4 เป็นการทดลองเมื่อใช้โครงข่ายนิรวลที่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญพบว่าช่วงของขนาดปริภูมิของการวิเคราะห์ส่วนประกอบमुखสำคัญที่นำมาใช้ที่ให้ผลที่ดีที่สุดสำหรับลักษณะเฉพาะฮิสโตแกรมสีอยู่ระหว่าง 20 ถึง 30 สำหรับ 4 ประเภทกีฬา ได้แก่ มวยไทย ฟุตบอล เทนนิส และ วอลเลย์บอล หลังจากนั้นแนวโน้มลดลงเมื่อขนาดปริภูมิที่นำมาใช้ใน

การจำแนกมีขนาดปริภูมิสูงขึ้นทั้งที่จำนวนหน่วยนิรอลในชั้นชอนมีจำนวนมากขึ้น และ เมื่อจำนวนหน่วยนิรอลในชั้นชอนมีจำนวน 12 หน่วยจะในช่วงขนาดของปริภูมิระหว่าง 20 ถึง 30 จะให้ผลความแม่นยำที่ดีที่สุดประมาณ ร้อยละ 86 ถึง 88 ดังรูปที่ 4.3

ดังนั้นสรุปได้ว่าในการการออกแบบโครงข่ายนิรอล สำหรับลักษณะเฉพาะฮิสโตแกรมสีไม่จำเป็นต้องใช้ขนาดของปริภูมิที่มีขนาดใหญ่จากการทดลองขนาดปริภูมิที่เหมาะสมสำหรับ 4 ประเภทกีฬาที่มีขนาดปริภูมิที่ 20 ถึง 30 มิติ ส่วนจำนวนหน่วยนิรอลจะมีร้อยละความแม่นยำมีแนวโน้มคงที่จึงไม่จำเป็นต้องสร้างโครงข่ายนิรอลที่มีขนาดใหญ่มากนัก จากผลการทดลองด้วยลักษณะเฉพาะฮิสโตแกรมสีจำนวนหน่วยนิรอลในชั้นชอนเป็น 28 หน่วย และ ขนาดปริภูมิของการวิเคราะห์ส่วนประกอบमुखสำคัญเป็น 24 มิติจะให้ผลความแม่นยำที่ดีที่สุดประมาณร้อยละ 87.30 ซึ่งแสดงคอนฟิวชันเมทริกซ์ของความแม่นยำที่ดีที่สุดแสดงได้ดังตารางที่ 4.2

ตารางที่ 4.2 คอนฟิวชันเมทริกซ์ของฮิสโตแกรมสีที่ให้ผลความแม่นยำที่ดีที่สุด

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา			
กีฬา	มวยไทย	ฟุตบอล	เทนนิส	วอลเลย์บอล
มวยไทย	90.13	2.66	3.33	3.86
ฟุตบอล	5.20	84.80	4.13	5.86
เทนนิส	13.00	3.00	81.53	2.46
วอลเลย์บอล	3.86	2.33	1.06	92.73

จากตารางที่ 4.2 พบว่าสำหรับฮิสโตแกรมสีกีฬาประเภท วอลเลย์บอลและมวยไทย ให้ผลการจำแนกได้ดีที่สุด สำหรับกีฬาเทนนิสพบว่าจำแนกผิดพลาดเป็นมวยไทยและวอลเลย์บอล เนื่องจาก ทั้งสองประเภทกีฬามีฉากที่มุ่งเน้นไปยังตัวนักกีฬามากจึงทำให้การจำแนกพิจารณาจากสีผิวจึงทำให้เกิดความผิดพลาด สำหรับกีฬาฟุตบอลจำแนกไปเป็นมวยไทยมากที่สุด จากฉากที่เน้นไปยังตัวผู้เล่นเช่นกัน

4.2 การทดลองที่ 2 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิดีโอทัศนกีฬาของลักษณะเฉพาะอโตคอร์รีโรแกรมโดยใช้โครงข่ายนิเวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

เนื่องจากลักษณะเฉพาะของสีที่ปรากฏในแต่ละประเภทกีฬาสามารถนำมาจำแนกประเภทกีฬาได้ ซึ่งในการทดลองที่ 1 เป็นลักษณะเฉพาะฮิสโตแกรมสีซึ่งเป็นอัตราส่วนของสีภายในภาพแต่ในการจำแนกแต่ละเลยคุณสมบัติเชิงพื้นที่ของภาพไป การทดลองนี้จึงนำลักษณะเฉพาะของสีที่สามารถใช้แทนข้อมูลเชิงพื้นที่ คือ ลักษณะเฉพาะของอโตคอร์รีโรแกรมเพื่อทดสอบว่าอโตคอร์รีโรแกรมมีคุณสมบัติในการจำแนกประเภทกีฬาได้หรือไม่

วัตถุประสงค์

เพื่อปรับปรุงประสิทธิภาพของการจำแนกประเภทวิดีโอทัศนกีฬาโดยการนำลักษณะเฉพาะสี ด้วยลักษณะเฉพาะอโตคอร์รีโรแกรม ซึ่งใช้โครงข่ายนิเวรอลแบบต่างๆ ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ปัจจัยกำหนดในการทดลอง

- โครงสร้างของโครงข่ายนิเวรอลซึ่งออกแบบไว้ใน การทดลองแบ่งได้เป็น 4 กรณี ได้แก่
 - กรณีที่ 1 แบบไม่ขึ้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 2 แบบไม่มีขึ้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 3 แบบมี 1 ชั้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 4 แบบมี 1 ชั้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
- อัตราการเรียนรู้ เป็น 0.0001 และเพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น สำหรับทั้ง 4 โครงข่ายนิเวรอลที่ได้ออกแบบไว้
- ประเภทกีฬาสำหรับการทดลองเป็น 4 ประเภทกีฬา ได้แก่ มวยไทย เทนนิส ฟุตบอล และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปตวันักกีฬา และ ฉากผู้ชมจำนวน 14 วิดีทัศน์ รวมเป็นจำนวน 14 วิดีทัศน์ คิดเป็นเวลาประมาณ 12 ชั่วโมง ซึ่งดึงเฟรมสำคัญทุกๆ หนึ่งเฟรมจาก 100 เฟรม จะได้ภาพในแต่ละ

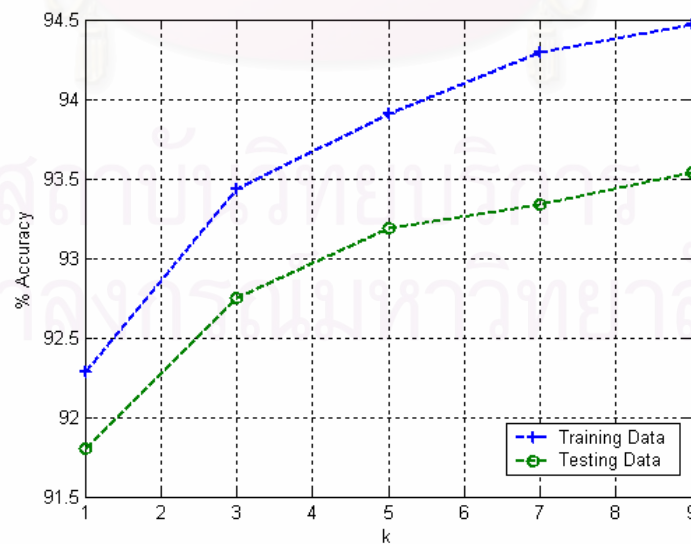
ประเภทกีฬาเป็น 1500 ภาพ รวมทั้งสิ้น 6000 ภาพ ภาพแต่ละภาพจะถูกลดขนาด จาก 384 x 448 จุดภาพ เป็น 96 x 112 จุดภาพและกำหนดเซตของระยะห่างเป็น $k = \{\{1\}, \{1,3\}, \{1,3,5\}, \{1,3,5,7\}, \{1,3,5,7,9\}\}$ ขนาดปริภูมิของเวกเตอร์ลักษณะเฉพาะ ออโตคอรัลโรแกรม เป็น 64 128 192 256 320 มิติ ตามลำดับ

- อัตราส่วนระหว่างจำนวนข้อมูลฝึกฝนต่อจำนวนข้อมูลทดสอบ เป็น 2 ต่อ 1
- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาเลดิเตด (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาเลดิเตด (Validate data) มีแนวโน้มสูงขึ้น
- ทดลองในแต่ละโครงข่ายนิเวรอลจำนวน 5 ครั้งแล้วจึงนำมาหาค่าเฉลี่ยของร้อยละความแม่นยำและคอนฟิวชันเมทริกซ์เฉลี่ย

ผลการทดลอง

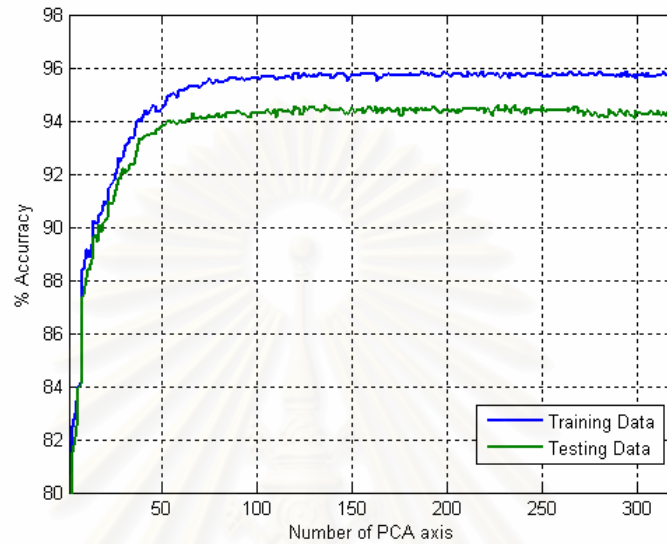
ผลการทดลองแบ่งออกเป็น 4 กรณีดังนี้

1. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ ซึ่งแสดงได้รูปที่ 4.5 เป็นเซตของระยะห่างเป็น $k = \{\{1\}, \{1,3\}, \{1,3,5\}, \{1,3,5,7\}, \{1,3,5,7,9\}\}$ ที่ใช้จำแนกประเภทกีฬา



- รูปที่ 4.5 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วมของข้อมูลทดสอบ สำหรับ ลักษณะเฉพาะออโตคอรัลโรแกรม

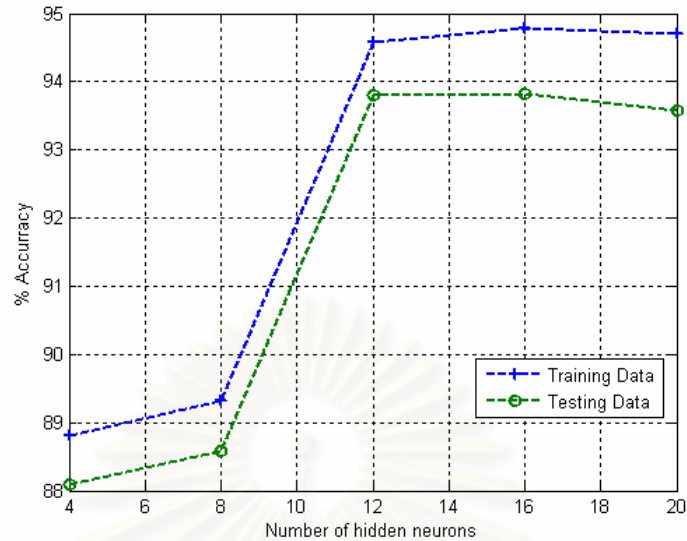
2. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ เมื่อเซตระยะห่างเป็น $k = \{1,3,5,7,9\}$ ด้วยการเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองดังรูปที่ 4.6



รูปที่ 4.6 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญร่วม สำหรับ ออโตคอรรีโลแกรม

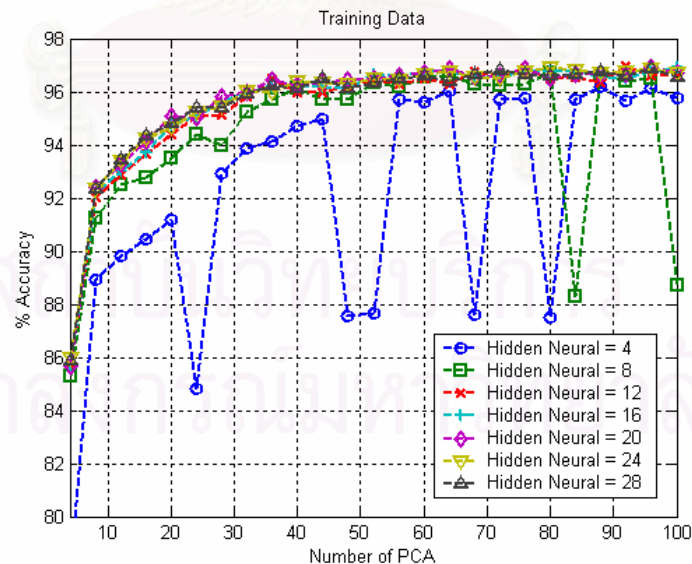
3. ผลการทดลองโครงข่ายที่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม เมื่อเซตระยะห่างเป็น $k = \{1,3,5,7,9\}$ โดยการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนได้ผลการทดลองดังรูปที่ 4.7

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

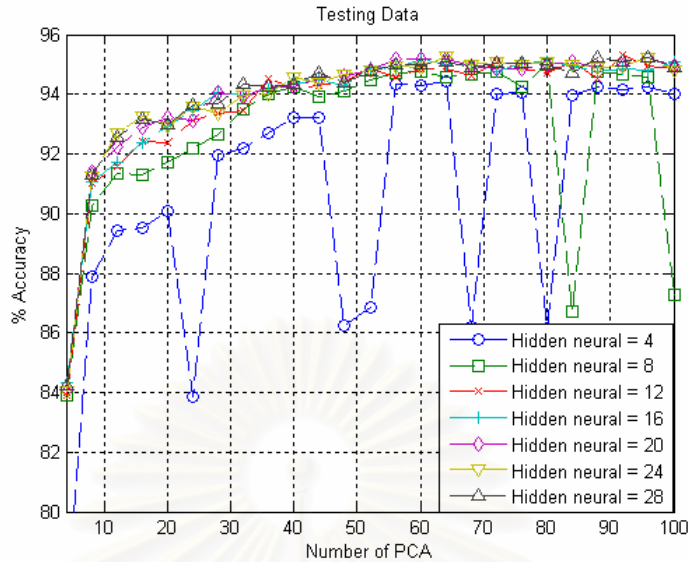


รูปที่ 4.7 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ออโตคอรวิไลแกรม

4. ผลการทดลองโครงข่ายนิวรอลชั้นซ่อน 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ เมื่อเซตระยะห่างเป็น $k = \{1,3,5,7,9\}$ ด้วยการเพิ่มขนาดปริภูมิส่วนประกอบमुखสำคัญ และ เพิ่มจำนวนหน่วยนิวรอลในชั้นซ่อน ซึ่งผลการทดลองได้ดังรูปที่ 4.8 และ รูปที่ 4.9 สำหรับชุดฝึกฝนและชุดทดสอบตามลำดับ



รูปที่ 4.8 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ออโตคอรวิไลแกรมของข้อมูลฝึกฝน



รูปที่ 4.9 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ออโตคอรรีโลแกรมของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลอง

จากการทดลองลักษณะเฉพาะของออโตคอรรีโลแกรม สำหรับกรณีที่ 1 ไม่มีชั้นซ่อน และไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วมนั้นเพื่อทดสอบหาเขตของระยะห่างที่เหมาะสมกับการจำแนกประเภทกีฬาผลการทดลองแสดงได้ดังรูปที่ 4.5 พบว่าเมื่อขนาดของเขตระยะห่างมีจำนวนมากขึ้นร้อยละความแม่นยำในการจำแนกประเภทกีฬาสูงขึ้นและมีแนวโน้มที่จะคงที่ ซึ่งการเพิ่มจำนวนของเขตระยะห่างนั้น ส่งผลให้ขนาดของปริภูมิออโตคอรรีโลแกรมก็จะมีขนาดใหญ่ตามจำนวนของเขตที่นำมาใช้ จากการทดลองดังรูปที่ 4.5 พบว่าเขตของระยะห่างที่ให้ผลความแม่นยำที่ดีที่สุดคือ $k = \{1,3,5,7,9\}$ ซึ่งจะให้ร้อยละความแม่นยำ สำหรับ ข้อมูลทดสอบที่ได้จากเขตของระยะห่างนี้ประมาณร้อยละ 93.60 เมื่อขนาดของปริภูมิเป็น 320 มิติ เมื่อนำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิวรอลที่ไม่มีชั้นซ่อน เมื่อเขตของระยะห่างที่นำมาใช้เป็น $k = \{1,3,5,7,9\}$ จะได้ผลการทดลองดังรูปที่ 4.6 พบว่าเมื่อขนาดของปริภูมิส่วนประกอบमुखสำคัญที่นำมาใช้มีขนาดปริภูมิมากขึ้น จะมีแนวโน้มร้อยละความแม่นยำสูงขึ้นจนกระทั่งมากกว่า 100 มิติ แล้วแนวโน้มของความแม่นยำจะคงที่ ซึ่งร้อยละความแม่นยำที่ได้จากข้อมูลทดสอบเป็นร้อยละ 94.50 ซึ่งให้ผลที่ดีกว่าไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญดังกรณีที่ 1

สำหรับกรณีที่ 3 เมื่อโครงสร้างของโครงข่ายนิวรอลที่ได้ออกแบบมีชั้นซ่อน 1 ชั้นซ่อน และทดลองปรับจำนวนหน่วยนิวรอลในชั้นซ่อนผลการทดลองได้ดังรูปที่ 4.7 พบว่าเมื่อจำนวนหน่วยนิวรอลเพิ่มมากขึ้นแนวโน้มของความแม่นยำจะสูงขึ้นจนกระทั่งมากกว่า 12 หน่วย

นิรวลความแม่นยำจะคงที่และมีแนวโน้มคงที่ เมื่อจำนวนหน่วยนิรวลมีจำนวนมากขึ้น สำหรับกรณีที่ 4 เมื่อโครงสร้างของโครงข่ายนิรวลมีชั้นซ่อน 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ ผลการทดลองของข้อมูลทดสอบแสดงได้ดังรูปที่ 4.9 พบว่าเมื่อเพิ่มขนาดของปริภูมิส่วนประกอบมุขสำคัญจะมีแนวโน้มสูงขึ้นในช่วงแรก และแนวโน้มคงที่เมื่อขนาดปริภูมิมากกว่า 60 มิติสำหรับทุกโครงสร้างของโครงข่ายนิรวล โดยที่เมื่อจำนวนหน่วยนิรวลในชั้นซ่อนมากกว่า 8 หน่วยนิรวลแนวโน้มของความแม่นยำของการจำแนกจะคงที่เป็นร้อยละ 95.2

ดังนั้นจากการทดลองสร้างโครงข่ายนิรวลทั้ง 4 กรณีพบว่าสำหรับลักษณะเฉพาะออโตคอร์รีโลแกรม เมื่อใช้การวิเคราะห์ส่วนประกอบมุขสำคัญร่วมกับโครงข่ายนิรวลแบบมีชั้นซ่อน 1 ชั้นซ่อน ให้ผลความแม่นยำที่ดีที่สุดซึ่งมากถึงร้อยละ 95.2 ณ ขนาดปริภูมิเป็น 60 มิติและจำนวนหน่วยนิรวลในชั้นซ่อนเป็น 20 หน่วยนิรวล ซึ่งแสดงคอนฟิวชันเมทริกซ์ได้ดังตารางที่ 4.3

ตารางที่ 4.3 คอนฟิวชันเมทริกซ์ของออโตคอร์รีโลแกรมที่ให้ผลความแม่นยำที่ดีที่สุด

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา			
กีฬา	มวยไทย	ฟุตบอล	เทนนิส	วอลเลย์บอล
มวยไทย	95.00	2.46	2.26	0.26
ฟุตบอล	2.46	94.46	2.80	0.26
เทนนิส	4.53	1.46	93.33	0.66
วอลเลย์บอล	0.80	0.13	1.06	98.00

จากตารางที่ 4.3 พบว่าออโตคอร์รีโลแกรมสามารถนำมาจำแนกได้ทุกประเภทกีฬาซึ่งความแม่นยำของการจำแนกประเภทกีฬาสำหรับทั้ง 4 ประเภทกีฬา ให้ผลความแม่นยำที่ดีกว่าวิธีอื่น เนื่องจากนำข้อมูลทางพื้นที่เข้ามาช่วยจำแนก ในฉากเน้นนักกีฬาจึงทำให้ความแม่นยำของการจำแนกมีค่าสูงขึ้น

4.3 การทดลองที่ 3 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิถีทัศนิกกีฬาของลักษณะเฉพาะการเคลื่อนที่ของกล้องโดยใช้โครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

จากแนวความคิดที่ว่ากีฬาแตกต่างประเภทกัน เกิดจากการบันทึกจากกล้องวิถีทัศนิกที่มีการเคลื่อนที่ที่ต่างต่างกัน ดังนั้นในการทดลองนี้เป็นการทดสอบการจำแนกประเภทวิถีทัศนิกกีฬา โดยใช้ลักษณะการเคลื่อนที่ของกล้องว่ามีความสามารถเป็นตัวแทนประเภทกีฬาได้หรือไม่

วัตถุประสงค์

เพื่อทดสอบประสิทธิภาพของการจำแนกประเภทวิถีทัศนิกกีฬาโดยการวิเคราะห์ลักษณะเฉพาะการเคลื่อนที่ของกล้อง ซึ่งใช้โครงข่ายนิวรอลแบบต่างๆ ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ปัจจัยกำหนดในการทดลอง

- โครงสร้างของโครงข่ายนิวรอลซึ่งออกแบบไว้ใน การทดลองแบ่งได้เป็น 4 กรณี ได้แก่
 - กรณีที่ 1 แบบไม่ขึ้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 2 แบบไม่มีขึ้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 3 แบบมี 1 ชั้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 4 แบบมี 1 ชั้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
- อัตราการเรียนรู้ เป็น 0.0001 และ เพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น สำหรับทั้ง 4 โครงข่ายนิวรอลที่ได้ออกแบบไว้
- ประเภทกีฬาสำหรับทดลองเป็น 4 ประเภทกีฬา ได้แก่ มวยไทย เทนนิส ฟุตบอล และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปตัวนักกีฬา และ ฉากผู้ชมจำนวน 14 วิดีทัศน์ รวมเป็นจำนวน 14 วิดีทัศน์ คิดเป็นเวลาประมาณ 12 ชั่วโมง กำหนดลักษณะการเคลื่อนที่ของกล้องให้แต่ละเฟรมของวีดีทัศน์ เป็น 8 รูปแบบคือ กล้องไม่มีการเคลื่อนที่ การเคลื่อนที่ของกล้องสายไปทางขวา (0 องศา) การเคลื่อนที่

ของกล่องสายไปทางซ้าย (180 องศา) การเคลื่อนที่ของกล่องก้ม (270 องศา) การเคลื่อนที่ของกล่องเงย (90 องศา) การเคลื่อนที่ของกล่องในแนว 45 องศา การเคลื่อนที่ของกล่องในแนว 135 องศา การเคลื่อนที่ของกล่องในแนว 225 องศา และการเคลื่อนที่ของกล่องในแนว 315 องศา โดยในกำหนดให้ทุกข้อออกจากลำดับวิธีทัศน์จะมีขนาด 100 เฟรม โดยกำหนดลักษณะการเคลื่อนที่ ทุก 5 เฟรม ประกอบด้วยทิศทาง (20 เมตร) และ ความเร็วของการเคลื่อนที่ของกล่อง (20 เมตร) ซึ่งจะได้เวกเตอร์ลักษณะเฉพาะการเคลื่อนที่ที่กล่องมีขนาดเป็น 40 เมตร

- อัตราส่วนระหว่างจำนวนข้อมูลฝึกฝนต่อจำนวนข้อมูลทดสอบ เป็น 2 ต่อ 1
- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาเลดิเตต (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาเลดิเตต (Validate data) มีแนวโน้มสูงขึ้น
- ทดลองในแต่ละโครงข่ายนิรอรลจำนวน 5 ครั้งแล้วจึงนำมาหาค่าเฉลี่ยของร้อยละความแม่นยำและคอนฟิวชันเมทริกซ์เฉลี่ย

ผลการทดลอง

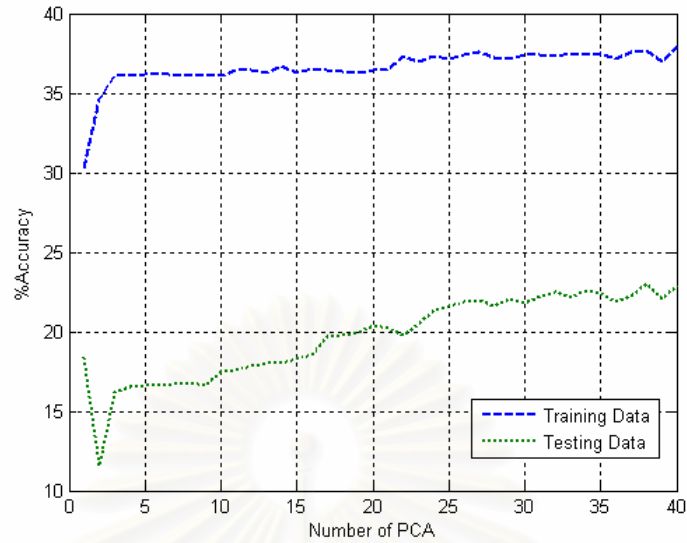
ผลการทดลองแบบออกเป็น 4 กรณีดังนี้

1. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ ซึ่งแสดงดังตารางที่ 4.4

ตารางที่ 4.4 ความแม่นยำของการจำแนกด้วยโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล่อง

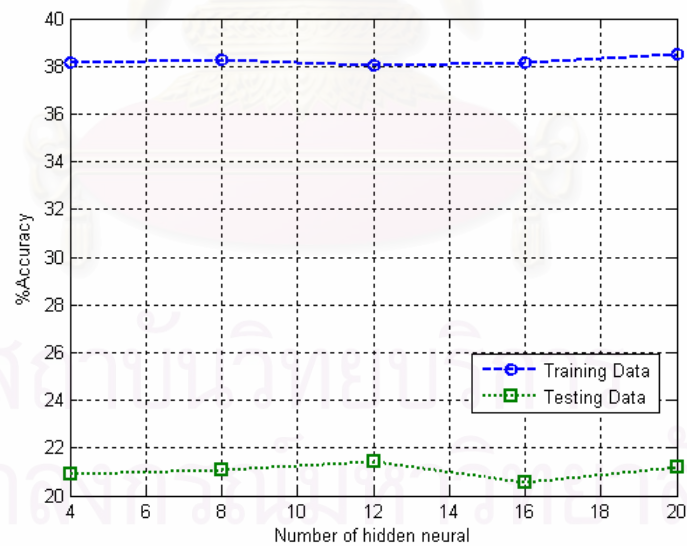
	ข้อมูลฝึกฝน	ข้อมูลทดสอบ
ร้อยละความแม่นยำ	37.41	22.74

2. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองดังรูปที่ 4.10



รูปที่ 4.10 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล้อง

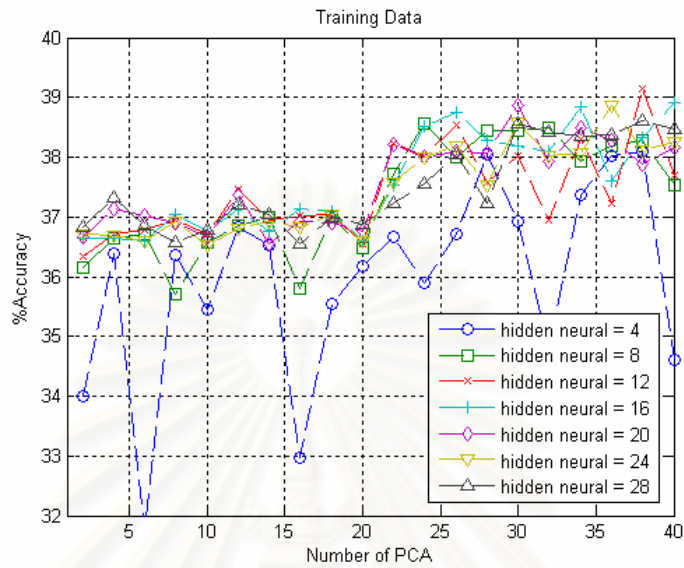
3. ผลการทดลองโครงข่ายที่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม โดยการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนได้ผลการทดลองดังรูปที่ 4.11



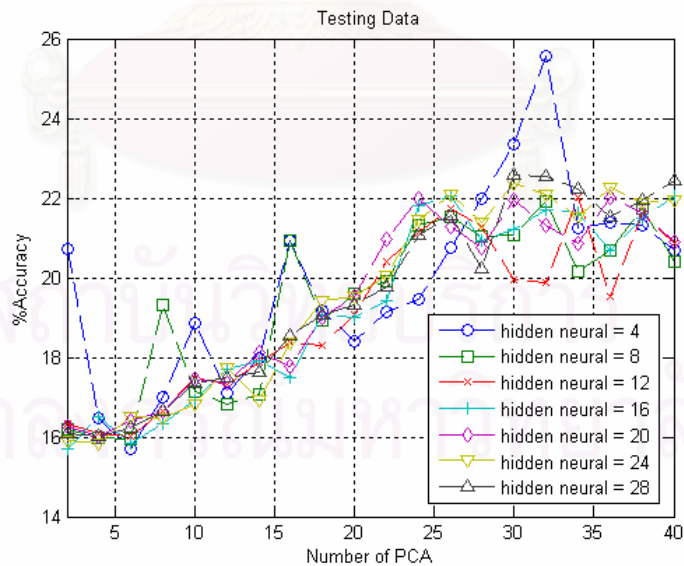
รูปที่ 4.11 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล้อง

4. ผลการทดลองโครงข่ายนิรวลชั้นซ่อน 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดปริภูมิส่วนประกอบमुखสำคัญ และ เพิ่มจำนวนหน่วย

นิเวรอลในชั้นซ่อน ซึ่งผลการทดสอบได้ดังรูปที่ 4.12 และ รูปที่ 4.13 สำหรับข้อมูลฝึกฝนและทดสอบ ตามลำดับ



รูปที่ 4.12 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล่องของข้อมูลฝึกฝน



รูปที่ 4.13 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล่องของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลอง

จากการทดลองโครงข่ายนิเวศทั้ง 4 โครงข่าย ลักษณะเฉพาะการเคลื่อนที่ของกิ้งก่า ไม่สามารถนำมาจำแนกประเภทกีฬาได้ แม้ว่าจะออกแบบให้โครงข่ายนิเวศมีชั้นซ้อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญแล้วก็ตาม เนื่องจากสำหรับ 4 ประเภทกีฬามีลักษณะการเคลื่อนที่ของกิ้งก่าที่คล้ายคลึงกัน ทำให้การเคลื่อนที่ของกิ้งก่าไม่สามารถเป็นตัวแทนสำหรับประเภทกีฬาได้ ซึ่งจากการทดลองกรณีนี้ 4 ผลการทดลองแสดงให้เห็นว่าความแม่นยำของการจำแนกประเภทกีฬาให้ผลที่ดีที่สุด เมื่อปริภูมิส่วนประกอบमुखสำคัญและจำนวนหน่วยนิเวศในชั้นซ้อนเป็น 34 มิติ และ 4 หน่วยนิเวศตามลำดับ ได้ร้อยละความแม่นยำเป็นร้อยละ 25.50 ซึ่งแสดงคอนฟิวชันเมตริกซ์ได้ดังตารางที่ 4.5

จากตารางที่ 4.5 ลักษณะเฉพาะการเคลื่อนที่ของกิ้งก่าจะจำแนกวิถีทัศน์ไปเป็นฟุตบอลเนื่องจากในกีฬาฟุตบอลเกิดการเคลื่อนที่ได้ทุกการเคลื่อนที่ โครงข่ายนิเวศจึงทำให้ความน่าจะเป็นของฟุตบอลมีความน่าจะเป็นสูงที่สุด

ตารางที่ 4.5 คอนฟิวชันเมตริกซ์ของการเคลื่อนที่ของกิ้งก่าที่ให้ผลความแม่นยำที่ดีที่สุด

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา			
กีฬา	มวยไทย	ฟุตบอล	เทนนิส	วอลเลย์บอล
มวยไทย	9.33	21.80	50.53	18.33
ฟุตบอล	22.73	1.00	40.93	35.33
เทนนิส	12.33	9.66	57.86	20.13
วอลเลย์บอล	17.20	16.40	32.40	34.00

4.4 การทดลองที่ 4 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิถีทัศน์กีฬาของลักษณะเฉพาะขอบสีโดยใช้โครงข่ายนิเวศร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

จากแนวความคิดที่ว่ากีฬาแตกต่างกัน รูปร่าง หรือ ลักษณะของขอบจะมีความแตกต่างกัน ดังนั้นในการทดลองนี้เป็นการทดสอบการจำแนกประเภทวิถีทัศน์กีฬา โดยใช้ลักษณะการขอบสี เพื่อพิสูจน์ว่ามีสามารถเป็นตัวแทนประเภทกีฬาได้หรือไม่

วัตถุประสงค์

เพื่อทดสอบประสิทธิภาพของการจำแนกประเภทวีดิทัศน์กีฬาโดยใช้ลักษณะเฉพาะของบอลี่ ซึ่งใช้โครงข่ายนิเวศแบบต่างๆ ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ปัจจัยกำหนดในการทดลอง

- โครงสร้างของโครงข่ายนิเวศซึ่งออกแบบไว้ใน การทดลองแบ่งได้เป็น 4 กรณี ได้แก่
 - กรณีที่ 1 แบบไม่ขึ้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 2 แบบไม่มีขึ้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 3 แบบมี 1 ชั้นซ้อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
 - กรณีที่ 4 แบบมี 1 ชั้นซ้อนและใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม
- อัตราการเรียนรู้ เป็น 0.0001 และเพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น สำหรับทั้ง 4 โครงข่ายนิเวศ
- ประเภทกีฬาสำหรับทดลองเป็น 4 ประเภทกีฬา ได้แก่ มวยไทย เทนนิส ฟุตบอล และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปต่อนักกีฬา และ ฉากผู้ชมจำนวน 14 วีดิทัศน์ รวมเป็นจำนวน 14 วีดิทัศน์ คิดเป็นเวลาประมาณ 12 ชั่วโมง ซึ่งดึงเฟรมสำคัญทุกๆ หนึ่งเฟรมจาก 100 เฟรม เฟรมสำคัญทุกภาพถูกลดขนาดจาก 384x448 จุดภาพ เป็น 96x112 จุดภาพ และ กำหนดความหมายของขอบในแต่ละบล็อกซึ่งความหมายของขอบแบ่งได้เป็น 5 ความหมาย คือ ขอบในแนวนอน (-) ขอบในแนวตั้ง (|) ขอบในแนว 45 องศา (/) และ ขอบในแนว 135 องศา (\) และ ไม่มีขอบ (null) ในแต่ละบล็อกถูกกำหนดความหมายขอบโดยใช้ฮิสโตแกรมของขอบในแต่ละบล็อกที่ให้จำนวนความหมายมากที่สุด เมื่อขนาดบล็อกที่ใช้เป็น 16x16 จุดภาพ ขนาดของเวกเตอร์ลักษณะเฉพาะบอลี่ที่ได้เป็น 126 มิติ
- อัตราส่วนระหว่างจำนวนข้อมูลฝึกฝนต่อจำนวนข้อมูลทดสอบเป็น 2 ต่อ 1

- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาลิเดต (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาลิเดต (Validate data) มีแนวโน้มสูงขึ้น
- ทดลองในแต่ละโครงข่ายนิรอรลจำนวน 5 ครั้งแล้วจึงนำมาหาค่าเฉลี่ยของร้อยละความแม่นยำและคอนฟิวชันเมทริกซ์เฉลี่ย

ผลการทดลอง

ผลการทดลองแบ่งออกเป็น 4 กรณีดังนี้

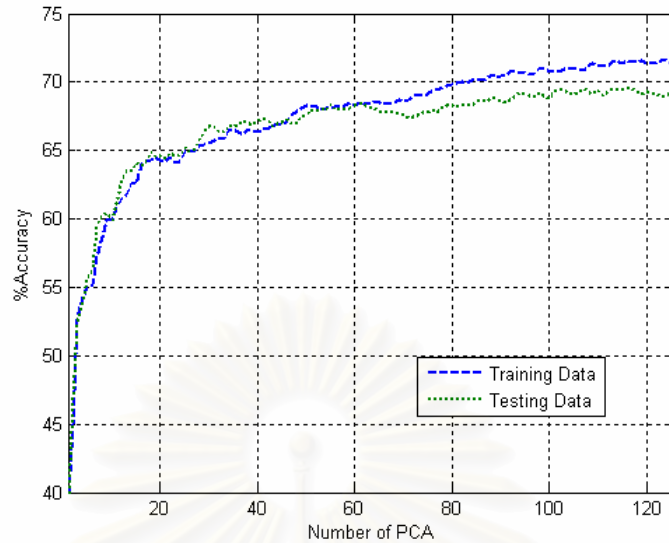
1. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ ซึ่งแสดงดังตารางที่ 4.6

ตารางที่ 4.6 ความแม่นยำของการจำแนกด้วยโครงข่ายที่ไม่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสี

	ข้อมูลฝึกฝน	ข้อมูลทดสอบ
ร้อยละความแม่นยำ	64.46	62.93

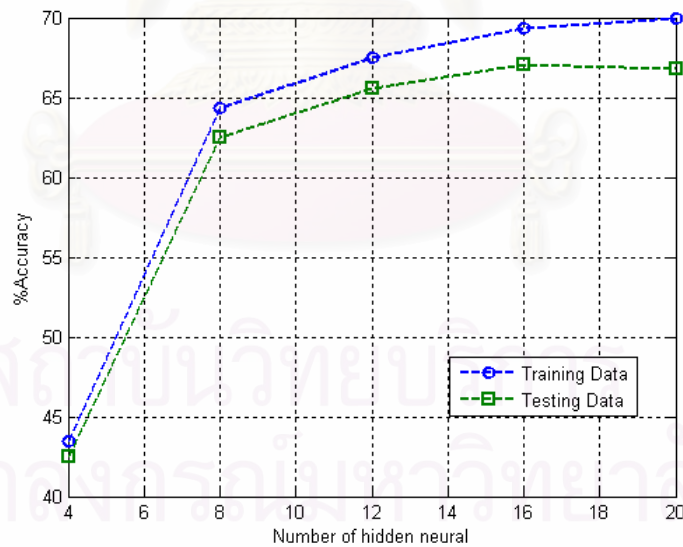
2. ผลการทดลองโครงข่ายที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองดังรูปที่ 4.14

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.14 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสี

3. ผลการทดลองโครงข่ายที่มีชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วม โดยการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนได้ผลการทดลองดังรูปที่ 4.15



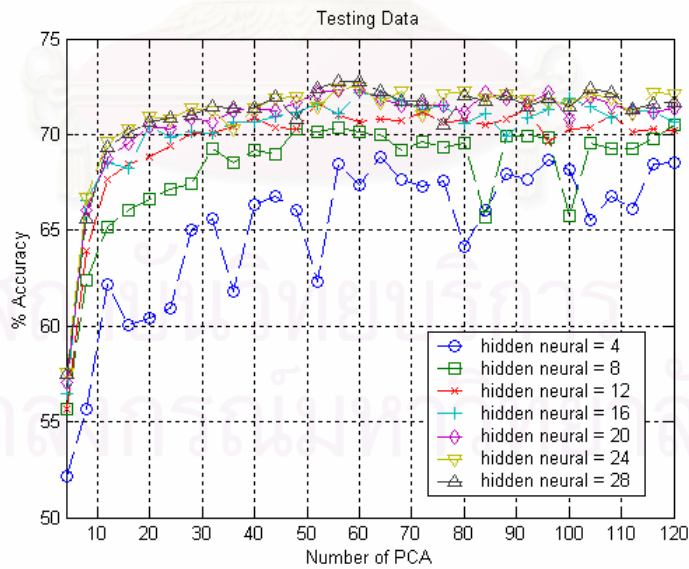
รูปที่ 4.15 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนและไม่ใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสี

4. ผลการทดลองโครงข่ายนิรวลชั้นซ่อน 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ ด้วยการเพิ่มขนาดปริภูมิส่วนประกอบमुखสำคัญ และ เพิ่มจำนวนหน่วย

นิเวรอลในชั้นซ่อน ซึ่งผลการทดสอบแสดงได้ดังรูปที่ 4.16 และ รูปที่ 4.17 สำหรับชุดฝึกฝนและชุดทดสอบ ตามลำดับ



รูปที่ 4.16 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน



รูปที่ 4.17 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบมี 1 ชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ ลักษณะเฉพาะขอบสีของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลอง

จากผลการทดลองกรณีที่ 1 ดังตารางที่ 4.6 ผลความแม่นยำของการจำแนกประเภทวิถีทัศนกีฬา 4 ประเภท มีความแม่นยำเป็นร้อยละ 62.93 สำหรับ ข้อมูลทดสอบ เมื่อเปรียบเทียบกับโครงข่ายนิรอรอลที่นำการวิเคราะห์ส่วนประกอบमुखสำคัญเข้าร่วม ซึ่งผลการทดลองแสดงดังรูปที่ 4.14 พบว่าขนาดของปริภูมิที่ในช่วงแรกความแม่นยำที่ได้จากข้อมูลทดสอบมีความแม่นยำสูงกว่าข้อมูลฝึกฝนแต่ยังไม่ใช้ผลความแม่นยำของการจำแนกที่ดีที่สุด ซึ่งผลความแม่นยำที่ดีที่สุดเมื่อขนาดปริภูมิมากกว่า 100 มิติขึ้นไป ซึ่งความแม่นยำอยู่ที่ประมาณร้อยละ 68.5

จากผลการทดลองในกรณีที่ 3 แสดงได้ดังรูปที่ 4.15 เมื่อออกแบบโครงข่ายนิรอรอลให้มี 1 ชั้นซ่อนพบว่าเมื่อจำนวนหน่วยนิรอรอลในชั้นซ่อนเพิ่มมากขึ้นจะทำให้ผลความแม่นยำของการจำแนกมีแนวโน้มสูงขึ้นและคงที่เมื่อจำนวนหน่วยนิรอรอลมีมากกว่า 16 หน่วยนิรอรอลซึ่งผลความแม่นยำที่ได้จากลักษณะเฉพาะของบอสี่ประมาณร้อยละ 67.5 และ เมื่อนำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิรอรอลแบบ 1 ชั้นซ่อน พบว่าสามารถได้ผลความแม่นยำของการจำแนกได้มากกว่าร้อยละ 70 ซึ่งผลความแม่นยำของการจำแนกที่ดีที่สุดในการ์ณที่ 4 เมื่อขนาดปริภูมิส่วนประกอบमुखสำคัญและจำนวนหน่วยนิรอรอลในชั้นซ่อนเป็น 56 มิติและ 28 หน่วยนิรอรอลตามลำดับ โดยคอนฟิวชันเมทริกซ์ของลักษณะเฉพาะของบอสี่ที่ดีที่สุดแสดงได้ดังตารางที่ 4.7 ซึ่งร้อยละของความแม่นยำที่ดีที่สุดเป็นร้อยละ 72.76 ของข้อมูลทดสอบ

ตารางที่ 4.7 คอนฟิวชันเมทริกซ์ของบอสี่ที่ให้ผลความแม่นยำที่ดีที่สุด

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา			
กีฬา	มวยไทย	ฟุตบอล	เทนนิส	วอลเลย์บอล
มวยไทย	68.60	11.60	11.33	8.46
ฟุตบอล	7.46	80.46	9.46	2.60
เทนนิส	7.20	10.86	69.86	12.06
วอลเลย์บอล	8.13	6.66	13.06	72.13

จากตารางที่ 4.7 พบว่าการจำแนกโดยใช้ขอบสำหรับกีฬาฟุตบอลให้ผลที่ดีที่สุด เนื่องจากในกีฬาฟุตบอลมีโครงสร้างของเส้นกรอบเห็นเด่นชัดที่สุดจึงสามารถทำให้จำแนกประเภทกีฬาได้ดีที่สุด

จากการทดลองที่ 1 ถึง 4 พบว่าเทคนิคที่นำเสนอโดยการใช้อาริวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิรอรสามารถปรับปรุงความแม่นยำในการจำแนกวิถีทัศน์ประเภทกีฬาได้ ซึ่งจะให้ผลดีกับลักษณะเฉพาะทุกลักษณะเฉพาะยกเว้น ลักษณะเฉพาะของการเคลื่อนที่ของกล้องเนื่องจาก ความเป็นไปได้ในการเคลื่อนที่ของกล้องในแต่ละประเภทกีฬามีโอกาสเคลื่อนที่ได้เท่าๆกัน ทำให้ความแปรปรวนในแต่ละแกนส่วนประกอบमुखสำคัญมีค่าสูง ทำให้โครงข่ายนิรอรไม่สามารถจำแนกประเภทกีฬาได้

4.5 การทดลองที่ 5 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิถีทัศน์กีฬาเมื่อจำนวนประเภทกีฬามีจำนวนมากขึ้นโดยใช้โครงข่ายนิรอรร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

เพื่อลดความซับซ้อนในกระบวนการฝึกฝนโครงข่ายนิรอรและดึงเอาเฉพาะข้อมูลที่มีนัยสำคัญ อีกทั้งช่วยลดขนาดปริภูมิข้อมูลตัวอย่างเพื่อให้มีประสิทธิภาพต่อการจำแนก จึงนำการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิรอร ซึ่งถือว่าการวิเคราะห์ส่วนประกอบमुखสำคัญเป็นกรรมวิธีประมวลผลเริ่มต้นให้กับข้อมูลตัวอย่าง ซึ่งในการทดลองนี้จะศึกษาถึงผลการกระทบเมื่อจำนวนประเภทวิถีทัศน์กีฬามีจำนวนมากขึ้น

วัตถุประสงค์

เพื่อทดสอบประสิทธิภาพของการจำแนกโดยโครงข่ายนิรอรร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสี ฮอโตคอรี่โรแกรม การเคลื่อนที่ของกล้อง และ ขอบสี เมื่อมีจำนวนประเภทกีฬามีจำนวนมากขึ้น

ปัจจัยกำหนดในการทดลอง

ในการทดลองนี้เป็นการทดสอบการวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิรอรที่ได้ผลความแม่นยำของการจำแนกที่ดีกับการจำแนก 4 ประเภทกีฬาจากการทดลองที่ 1 ถึง 4 ซึ่งในการทดลองที่ 5 นี้เป็นการศึกษาถึงผลกระทบเมื่อจำนวนประเภทกีฬามีจำนวนมากขึ้น สำหรับ 4 ลักษณะเฉพาะ ซึ่งในการทดลองนี้แบ่งเป็นหัวข้อย่อย 4 ส่วนตามลักษณะเฉพาะที่ได้นำเสนอในวิทยานิพนธ์เอบทที่ 3 ซึ่งมีปัจจัยกำหนดทั้ง 4 ลักษณะเฉพาะดังนี้

- โครงสร้างของโครงข่ายนิวรอลซึ่งออกแบบไว้เป็น 2 รูปแบบ ได้แก่
 - กรณีที่ 1 แบบไม่มีชั้นซ่อนรวมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ
 - กรณีที่ 2 แบบมี 1 ชั้นซ่อนรวมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ
- อัตราการเรียนรู้ เป็น 0.0001 และ เพิ่มอัตราการเรียนรู้เป็น 1.005 เท่าในแต่ละ epoch ที่เพิ่มขึ้น สำหรับทั้ง 2 โครงข่ายนิวรอล
- ประเภทกีฬาสำหรับทดลองเป็น 7 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด เทนนิส และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปตัวนักกีฬา และ ฉากผู้ชมจำนวน 27 วิดีทัศน์ คิดเป็นเวลาประมาณ 233 นาที เมื่อกำหนดให้ 1 ซีตประกอบด้วย 50 เฟรม ซึ่งจำนวนข้อมูลในการฝึกฝนและทดสอบแสดงได้ดังตารางที่ 4.8
- จำนวนประเภทกีฬาที่เพิ่มขึ้นมีลำดับดังต่อไปนี้
 - จำนวน 2 ประเภทกีฬา ได้แก่ บาสเก็ตบอล และ มวยไทย
 - จำนวน 3 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย และ ฟุตบอล
 - จำนวน 4 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล และ กอล์ฟ
 - จำนวน 5 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ และ สปริงบอร์ด
 - จำนวน 6 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด และ เทนนิส
 - จำนวน 7 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด เทนนิส และ วอลเลย์บอล
- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาลิเดต (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาลิเดต (Validate data) มีแนวโน้มสูงขึ้น

ตารางที่ 4.8 จำนวนข้อมูลฝึกฝนและข้อมูลทดสอบ สำหรับ 7 ประเภทกีฬา

	จำนวนวีดิทัศน์ (วีดิทัศน์)	จำนวนข้อมูลฝึกฝน (ชื่อวีดิทัศน์)	จำนวนข้อมูลทดสอบ (ชื่อวีดิทัศน์)
บาสเก็ตบอล	6	667	333
มวยไทย	3	667	333
ฟุตบอล	6	667	333
กอล์ฟ	4	667	333
สปริงบอร์ด	2	667	333
เทนนิส	3	667	333
วอลเลย์บอล	3	667	333
รวม	27	4669	2331

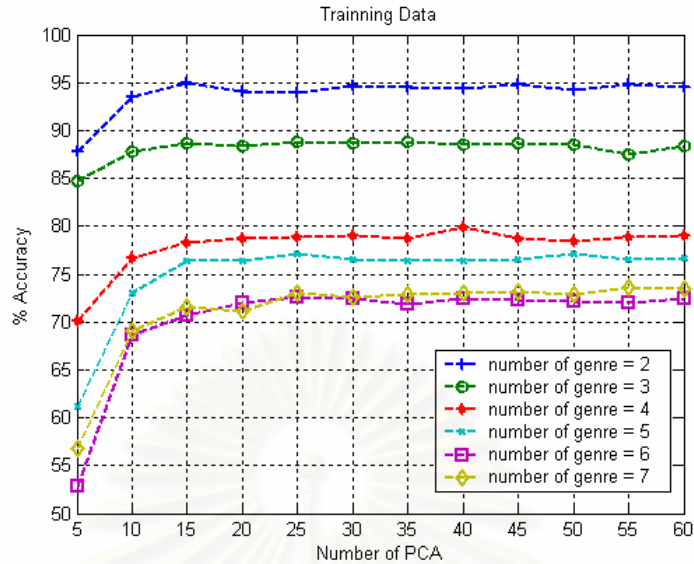
- ทดลองในแต่ละโครงข่ายนิรลจำนวน 3 ครั้งแล้วจึงนำมาหาค่าเฉลี่ยของร้อยละความแม่นยำและคอนฟิรชันเมทริกซ์เฉลี่ย

ผลการทดลอง

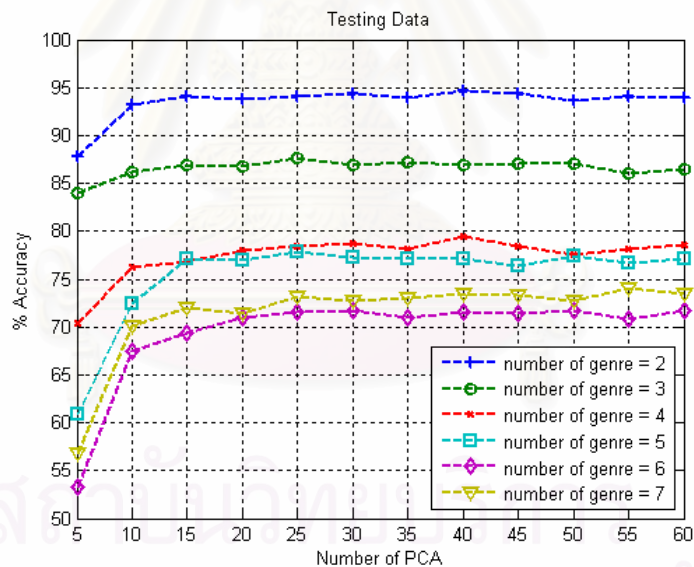
ซึ่งในการทดลองนี้แบ่งได้ผลการทดลอง 4 ผลการทดลองย่อยตามลักษณะเฉพาะดังต่อไปนี้

4.5.1 ผลการทดลองการจำแนกประเภทวีดิทัศน์กีฬาด้วยการใช้ฮิสโตแกรมสีโดยโครงข่ายนิรลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ผลการทดลองลักษณะเฉพาะฮิสโตแกรมสีขนาด 64 ถังสี จำแนกด้วยการใช้โครงข่ายนิรลที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ เมื่อเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.18 และ รูปที่ 4.19 ตามลำดับ

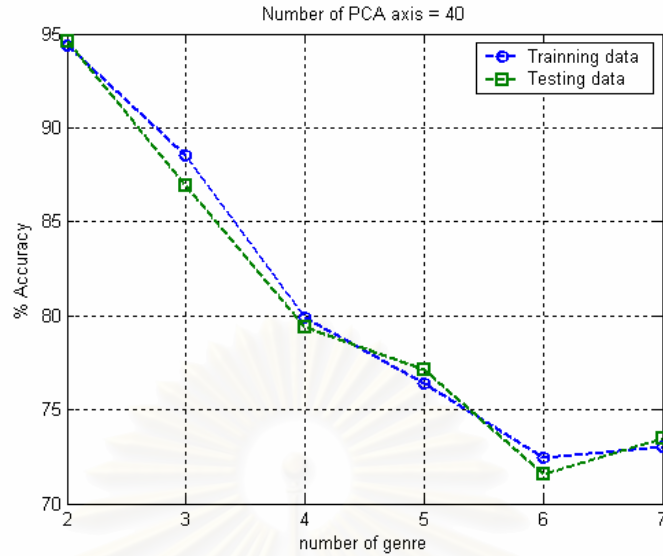


รูปที่ 4.18 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนรวมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน



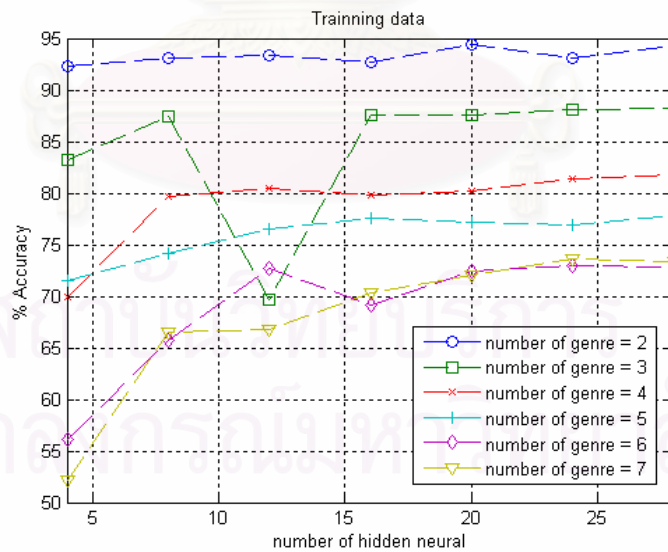
รูปที่ 4.19 ความแม่นยำของการจำแนกด้วยโครงข่ายนิวรอลแบบไม่มีชั้นซ่อนรวมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ

จากผลการทดลองดังรูปที่ 4.18 และ รูปที่ 4.19 ขนาดของปริภูมิที่เหมาะสมในการจำแนกหาได้จากขนาดปริภูมิที่ให้ค่าเฉลี่ยของความแม่นยำในการจำแนกทั้ง 7 ประเภทกีฬาที่สูงที่สุด ซึ่งลักษณะเฉพาะฮิสโตแกรม ได้ให้ผลความแม่นยำที่ดีที่สุดเมื่อปริภูมิมีขนาดเป็น 40 มิติ ซึ่งแสดงผลกระทบเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นได้รูปที่ 4.20

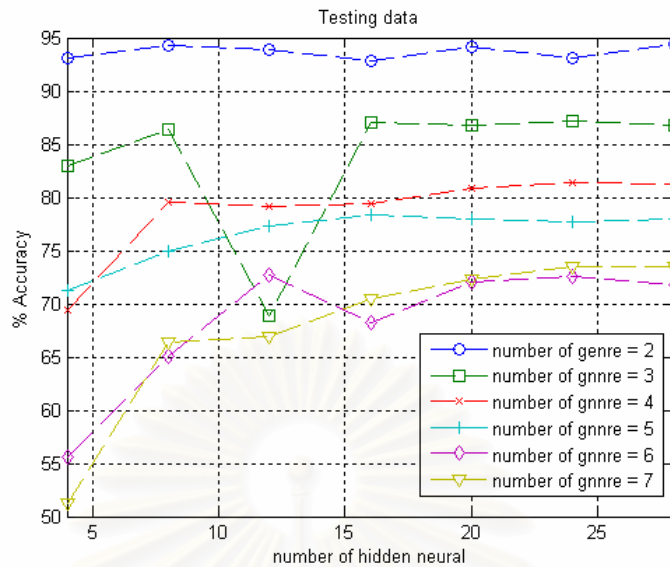


รูปที่ 4.20 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยใช้โครงข่ายนิรวลแบบไม่มีชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสี

ในการออกแบบโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนจะกำหนดให้จำนวนอินพุตของโครงข่ายเป็น 40 อินพุต ซึ่งได้ผลกระทบของจำนวนประเภทกีฬาที่เพิ่มขึ้นเทียบการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนสำหรับข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.21 และ รูปที่ 4.22 ตามลำดับ



รูปที่ 4.21 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยใช้โครงข่ายนิรวลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน



รูปที่ 4.22 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองหัวข้อย่อย 4.5.1

จากการทดลองโครงข่ายนิเวศแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญรูปที่ 4.18 และ รูปที่ 4.19 พบว่าการเพิ่มขนาดปริภูมิส่วนประกอบ मुखสำคัญจะไม่ช่วยในการเพิ่มความแม่นยำของการจำแนกประเภทกีฬา เมื่อจำนวนประเภทกีฬา มีจำนวนประเภทมากขึ้น โดยขนาดปริภูมิของฮิสโตแกรมสีเป็น 40 มิติที่ทำให้ผลความแม่นยำที่ดีที่สุดสำหรับโครงข่ายนิเวศแบบไม่มีชั้นซ่อนของทุกประเภทกีฬา ซึ่งเปรียบระหว่างข้อมูลฝึกฝน และข้อมูลทดสอบได้ดังรูปที่ 4.20 พบมาเมื่อจำนวนกีฬาเพิ่มมากขึ้นจะทำให้ความแม่นยำของการ จำแนกต่ำลงเนื่องจากข้อมูลตัวอย่างมีประเภทกีฬาเป็นจำนวนมากขึ้นทำให้ความซับซ้อนในการ ตัดสินใจมีมากขึ้นจึงทำให้ผลความแม่นยำต่ำลง แต่จากผลการทดลองเมื่อเพิ่มประเภทกีฬาเป็น 7 ประเภทมีแนวโน้มคงที่เนื่องจากอัตราส่วนของสีที่ใช้ในวอลเลย์บอลมีความแตกต่างจากประเภท กีฬาอื่นๆ อย่างชัดเจน จึงทำให้ผลที่ได้ได้เท่ากับ 6 ประเภทกีฬา เนื่องจากประเภทกีฬาทำให้ ข้อมูลที่ได้จากฮิสโตแกรมมีความซับซ้อนสูงจึงออกแบบโครงข่ายนิเวศแบบมีชั้นซ่อน 1 ชั้นเพื่อ ปรับปรุงประสิทธิภาพของโครงข่ายนิเวศร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญได้ผลการ ทดลองดังรูปที่ 4.21 และ รูปที่ 4.22 ของข้อมูลฝึกฝนและทดสอบ ตามลำดับ พบว่าเมื่อจำนวน หน่วยนิเวศในชั้นซ่อนที่ใช้จำเป็นต้องมีจำนวนมากขึ้น แต่ผลความแม่นยำเมื่อเปรียบเทียบกับ รูปที่ 4.20 ที่แสดงในกรณีไม่มีชั้นซ่อน ความแม่นยำจากชั้นซ่อนสูงขึ้นเพียงร้อยละ 1 ถึง 2 เท่านั้น สำหรับฮิสโตแกรมสี ซึ่งผลที่ดีที่สุดเมื่อเลือกใช้ขนาดปริภูมิการวิเคราะห์ส่วนประกอบमुखสำคัญ

เป็น 40 มิติเมื่อจำนวนหน่วยนิรวัลในชั้นซ้อนเป็น 28 หน่วยนิรวัลได้คอนฟิวชันเมทริกซ์ใน 7 ประเภทกีฬาตารางที่ 4.9

จากตารางที่ 4.9 พบว่าฮิสโตแกรมสีจำแนกโดยใช้โครงข่ายนิรวัลแบบมี 1 ชั้นซ้อนให้ผลการจำแนกในชนิดกีฬามวยไทยดีที่สุด ซึ่งสังเกตได้ว่าระหว่างฟุตบอลและกอล์ฟที่มีสีเหมือนกันโครงข่ายนิรวัลไม่สามารถแยกแยะกีฬาทั้งสองประเภทนี้ได้ เนื่องจากสีที่เหมือนกันทำให้ฮิสโตแกรมจึงมีแบบรูปคล้ายคลึงกัน และระหว่างกีฬาสปริงบอร์ดกับเทนนิส สีสนามของเทนนิสที่ใช้ในการทดลองเป็นสีฟ้าซึ่งจะไปเหมือนกับสีของน้ำในกีฬาสปริงบอร์ดจึงทำให้ผลการจำแนกมีความผิดพลาด

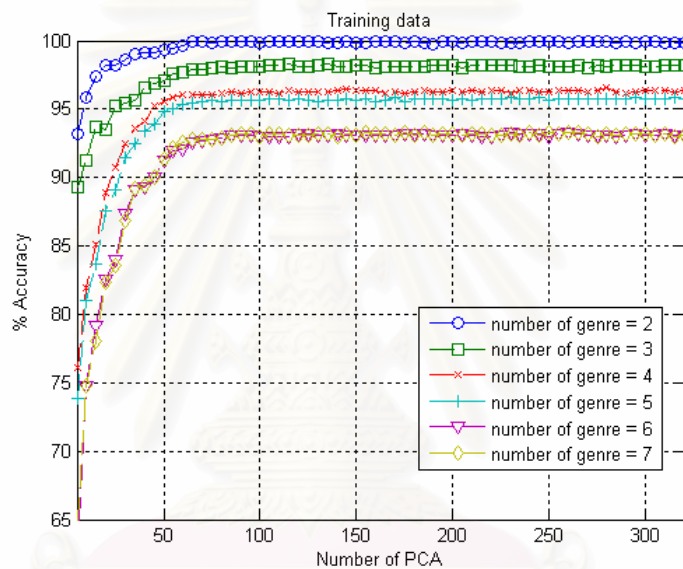
สรุปได้ว่าถ้าเพิ่มประเภทกีฬาจำนวนมากขึ้น เนื่องจากคุณสมบัติของฮิสโตแกรมสีที่รูปแบบที่เหมือนกันในทางสีจึงทำให้ไม่สามารถจำแนกกีฬาได้อย่างแม่นยำ

ตารางที่ 4.9 คอนฟิวชันเมทริกซ์ของฮิสโตแกรมสำหรับ 7 ประเภทกีฬา

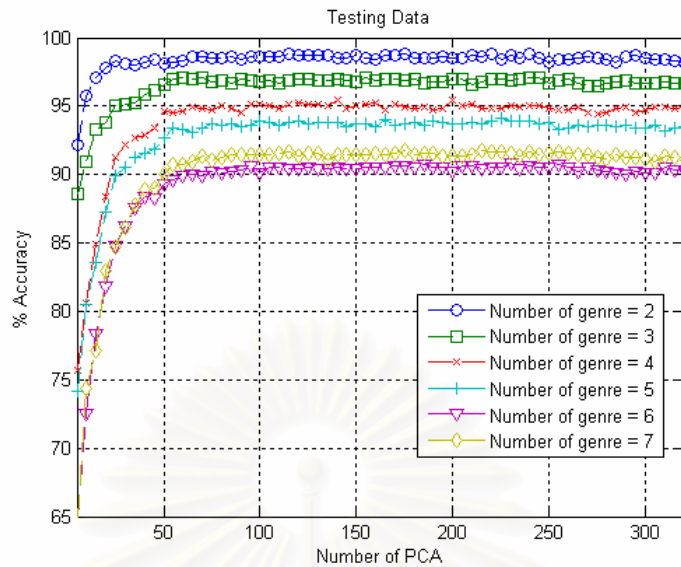
	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	76.57	9.81	4.70	1.00	1.00	5.50	1.40
มวยไทย	2.90	93.19	2.10	0.00	0.20	0.70	0.90
ฟุตบอล	6.10	6.90	59.75	17.61	1.60	3.90	4.10
กอล์ฟ	1.90	5.20	20.82	68.06	1.20	1.30	1.50
สปริงบอร์ด	1.90	3.40	11.01	3.00	52.55	21.12	7.00
เทนนิส	3.80	12.31	0.70	1.00	6.70	74.07	1.40
วอลเลย์บอล	2.00	3.60	0.20	0.00	2.70	1.40	90.09

4.5.2 ผลการทดลองการจำแนกประเภทวิดีโอทัศนกีฬาด้วยการใช้อัตนอร์รีไลแกรมโดยโครงข่ายนิเวรอลร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญ

ผลการทดลองลักษณะเฉพาะอตนอร์รีไลแกรม ซึ่งภาพแต่ละภาพจะถูกลดขนาดจาก 384 x 448 จุดภาพ เป็น 96 x 112 จุดภาพและกำหนดเซตของระยะห่างเป็น $k = \{1,3,5,7,9\}$ ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 320 มิติ จำแนกด้วยการใช้โครงข่ายนิเวรอลที่ไม่มีชั้นซ่อนร่วมกับการใช้การวิเคราะห์ส่วนประกอบมุขสำคัญ เมื่อเพิ่มขนาดของปริภูมิส่วนประกอบมุขสำคัญได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.23 และ รูปที่ 4.24 ตามลำดับ

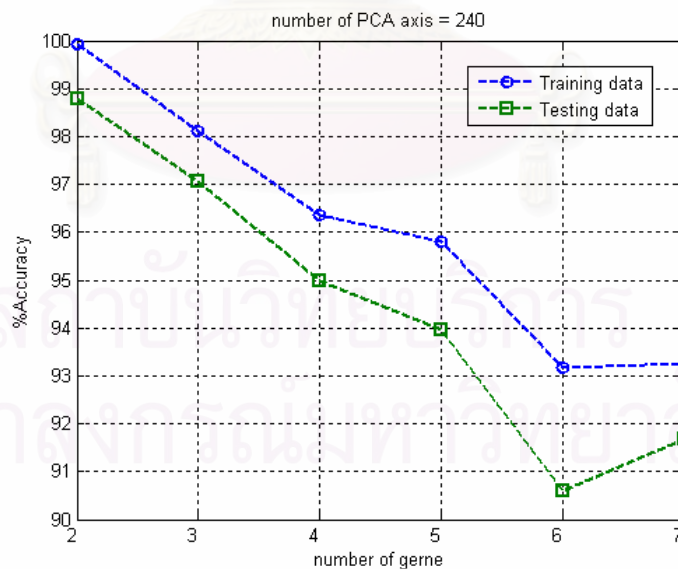


รูปที่ 4.23 ความแม่นยำของการจำแนกด้วยโครงข่ายนิเวรอลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบมุขสำคัญร่วม สำหรับ อตนอร์รีไลแกรมของข้อมูลฝึกฝน



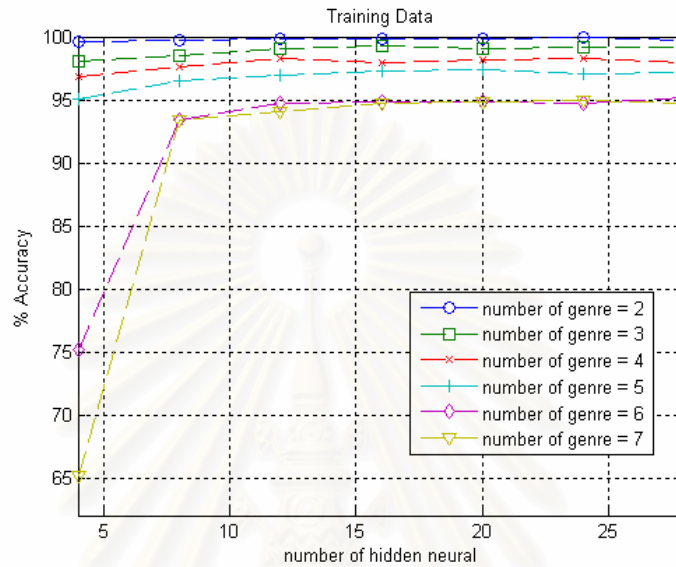
รูปที่ 4.24 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรพลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญร่วม สำหรับ ออโตคอรัลโแกรมของข้อมูลทดสอบ

จากผลการทดลองดังรูปที่ 4.23 และ รูปที่ 4.24 ขนาดของปริภูมิที่เหมาะสมในการจำแนกหาได้จากขนาดปริภูมิที่ให้ค่าเฉลี่ยของความแม่นยำในการจำแนกทั้ง 7 ประเภทกีฬาที่สูงที่สุด ซึ่งลักษณะเฉพาะออโตคอรัลโแกรม ได้ให้ผลความแม่นยำที่ดีที่สุดเมื่อปริภูมิมีขนาดเป็น 240 มิติ ซึ่งแสดงผลกระทบเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นได้ดังรูปที่ 4.25

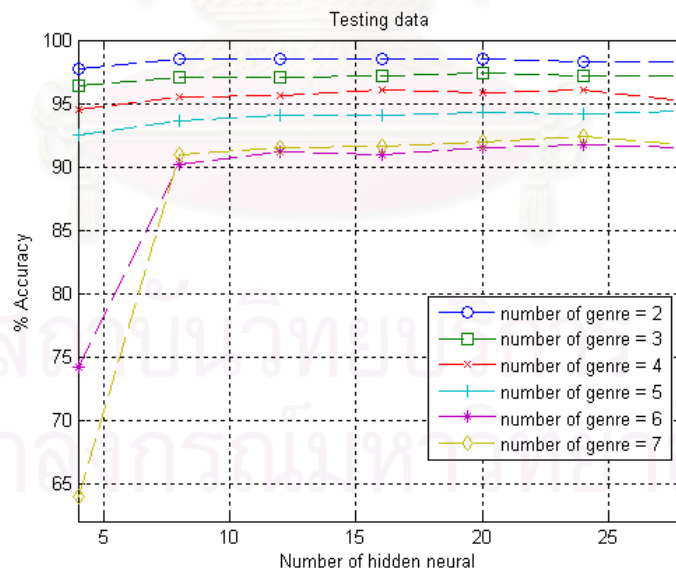


รูปที่ 4.25 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใชโครงข่ายนิรพลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญของออโตคอรัลโแกรม

ในการออกแบบโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนสำหรับอโตคอรี่โดแกรมจะกำหนดให้จำนวนอินพุตของโครงข่ายเป็น 240 อินพุต ซึ่งได้ผลกระทบของจำนวนประเภทกีฬาที่เพิ่มขึ้นเทียบการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนสำหรับข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.26 และ รูปที่ 4.27 ตามลำดับ



รูปที่ 4.26 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิรวลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ อโตคอรี่โดแกรมของข้อมูลฝึกฝน



รูปที่ 4.27 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิรวลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ อโตคอรี่โดแกรมของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองห้วยย่อย 4.5.2

จากการทดลองลักษณะเฉพาะอโตคอร์รีโลแกรมโดยใช้โครงข่ายนิวรรอลแบบไม่มีชั้นซ้อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญดังรูปที่ 4.23 และ รูปที่ 4.24 พบว่าการเพิ่มขนาดปริภูมิमुखสำคัญมากขึ้นจะทำให้ความแม่นยำของการจำแนกสูงขึ้น และมีแนวโน้มจะคงที่เมื่อขนาดปริภูมิมากกว่า 100 มิติ เมื่อจำนวนประเภทกีฬาเพิ่มขึ้น แม้ว่าขนาดของปริภูมิส่วนประกอบमुखสำคัญเพิ่มขึ้น แต่ความแม่นยำของการจำแนกต่ำลงเนื่องจากปริภูมิที่เพิ่มขึ้นปราศจากข้อมูลที่เป็นประโยชน์ต่อการจำแนกประเภทกีฬา จากการทดลองพบว่าเมื่อเพิ่มประเภทกีฬาที่แตกต่างทางด้านข้อมูลสี่อย่างชัดเจนดังการเพิ่มจาก 6 ประเภท เป็น 7 ประเภทที่มีเพิ่มกีฬาวอลเลย์บอลเข้าทำให้ผลความแม่นยำสูงขึ้นได้ ดังนั้นการการเพิ่มประเภทกีฬาที่มีความแตกต่างกันของสี่อย่างชัดเจนจะช่วยให้ผลความแม่นยำดีขึ้นได้ ซึ่งจากการทดลองพบว่าปริภูมิขนาด 240 มิติให้ผลการจำแนกดีที่สุดกับทุกการเพิ่มของประเภทกีฬาที่ได้ผลดังรูปที่ 4.25 เพื่อปรับปรุงความแม่นยำของการจำแนกให้สูงขึ้นจึงออกแบบโครงข่ายนิวรรอลแบบมี 1 ชั้นซ้อนโดยให้จำนวนอินพุตเท่ากับ 240 มิติ ซึ่งได้ผลการทดลองดังรูปที่ 4.26 และ รูปที่ 4.27 สำหรับข้อมูลทดสอบและฝึกฝนตามลำดับ พบว่าถ้าต้องการให้ผลความแม่นยำสูงขึ้นเมื่อจำนวนประเภทกีฬามากขึ้นต้องใช้จำนวนหน่วยนิวรรอลในชั้นซ้อนมากขึ้น แต่เมื่อให้จำนวนหน่วยนิวรรอลมากขึ้นจะทำให้ผลความแม่นยำของการจำแนกคงที่ ซึ่งผลที่ดีที่สุดสำหรับ 7 ประเภทกีฬาเมื่อขนาดปริภูมิเป็น 240 มิติ และ จำนวนหน่วยนิวรรอลในชั้นซ้อนเป็น 24 หน่วยนิวรรอล โดยที่แสดงคอนฟิวชันเมทริกซ์ได้ดังตารางที่ 4.10

จากตารางที่ 4.10 พบว่าอโตคอร์รีโลแกรมโดยใช้โครงข่ายนิวรรอลแบบ 1 ชั้นซ้อนให้ผลการจำแนกได้ทุกประเภทกีฬา ซึ่งสามารถแยกได้ทั้งฉากเน้นนักกีฬาและฉากผู้ชม

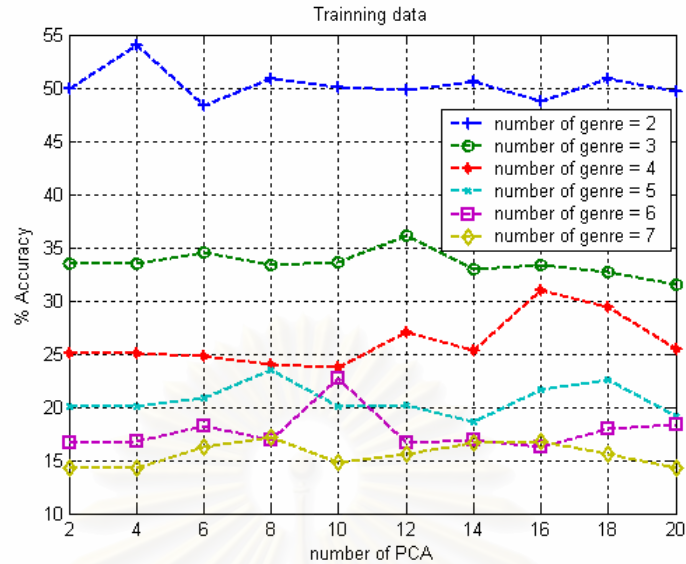
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.10 คอนฟิวชันเมทริกซ์ของอโตคอร์รีโลแกรมสำหรับ 7 ประเภทกีฬา

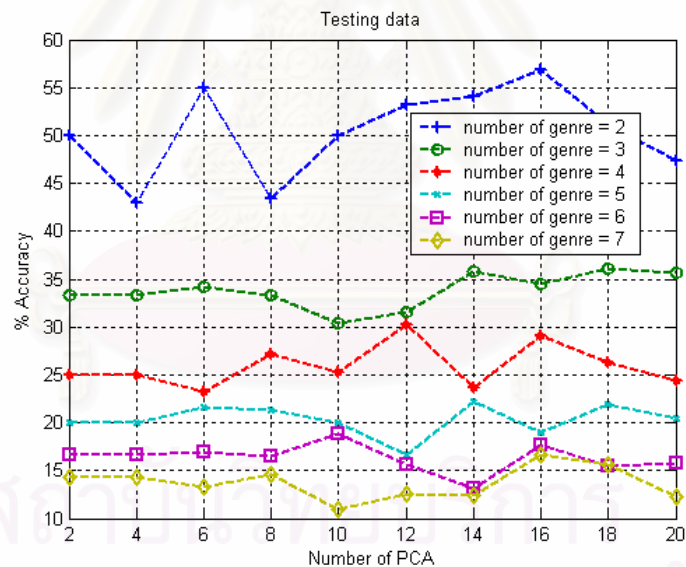
	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	92.39	1.60	3.60	0.50	0.30	1.10	0.50
มวยไทย	1.10	97.79	0.50	0.00	0.10	0.20	0.30
ฟุตบอล	2.70	0.80	87.38	3.40	2.80	1.90	1.10
กอล์ฟ	0.00	0.40	2.20	94.29	2.40	0.70	0.00
สปริงบอร์ด	0.00	0.00	1.40	1.80	93.29	2.90	0.60
เทนนิส	2.80	3.40	2.10	0.40	4.50	84.58	2.20
วอลเลย์บอล	0.40	0.30	0.00	0.30	1.00	0.70	97.29

4.5.3 ผลการทดลองการจำแนกประเภทวิถีทัศนกีฬาด้วยการใช้การเคลื่อนที่ของกล้อง โดยโครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ผลการทดลองลักษณะเฉพาะการเคลื่อนที่ของกล้อง ซึ่งดึงลักษณะการเคลื่อนที่ของกล้องทุก 5 เฟรม ใน 1 ซ็อต ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 20 มิติจำแนกด้วยการใช้โครงข่ายนิวรอลที่ไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ เมื่อเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.28 และ รูปที่ 4.29 ตามลำดับ

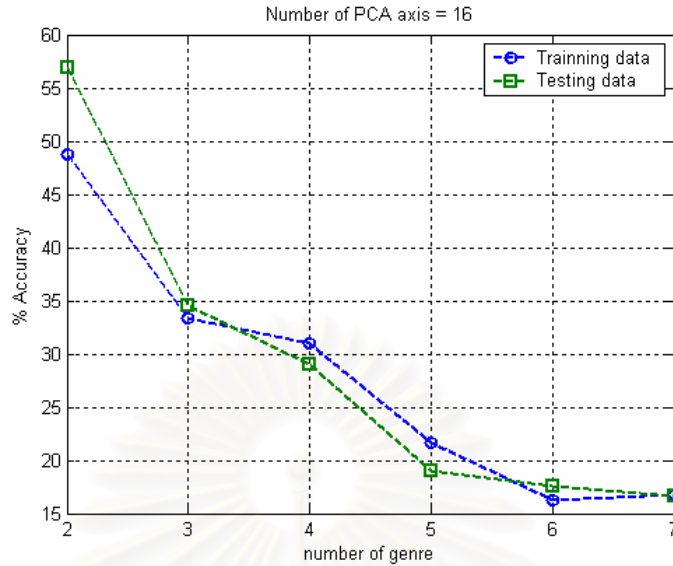


รูปที่ 4.28 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรพลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญร่วม สำหรับ การเคลื่อนที่ของกล่องของข้อมูลฝึกฝน



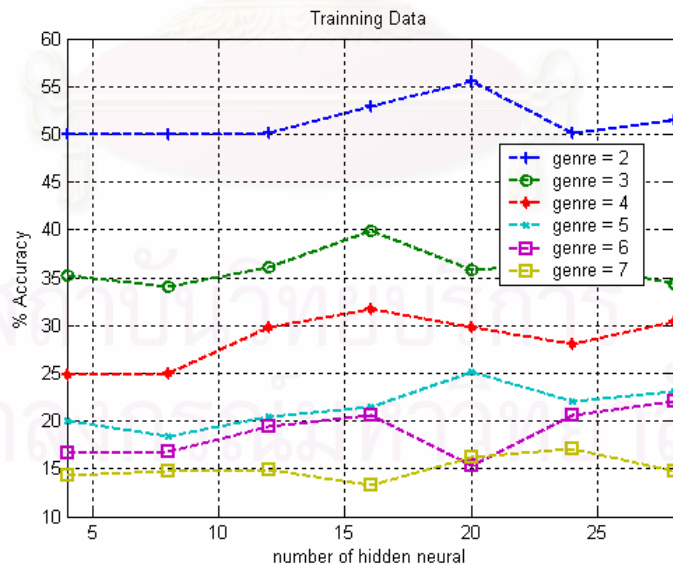
รูปที่ 4.29 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรพลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ การใช้การเคลื่อนที่ของกล่องของข้อมูลทดสอบ

จากผลการทดลองดังรูปที่ 4.28 และ รูปที่ 4.29 ขนาดของปริภูมิที่เหมาะสมในการจำแนกหาได้จากขนาดปริภูมิที่ให้ค่าเฉลี่ยของความแม่นยำในการจำแนกทั้ง 7 ประเภทกีฬาที่สูงที่สุด ซึ่งลักษณะเฉพาะการเคลื่อนที่ของกล่องให้ผลความแม่นยำที่ดีที่สุดเมื่อปริภูมิมีขนาดเป็น 16 มิติ ซึ่งแสดงผลกระทบเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นได้ดังรูปที่ 4.30

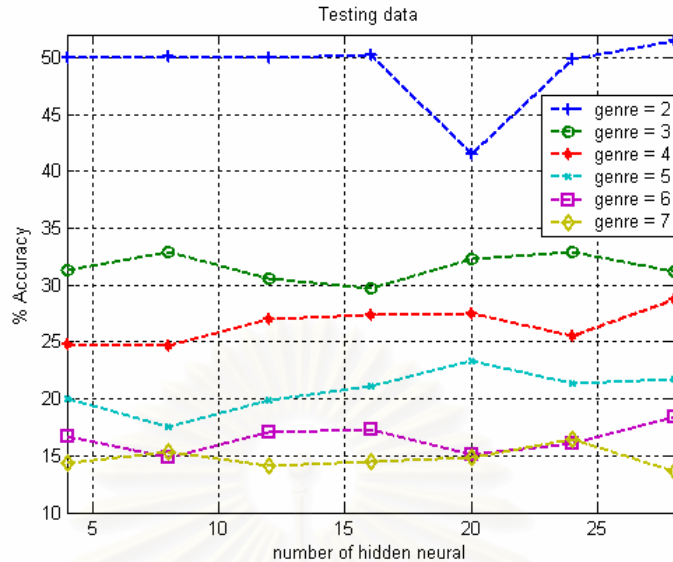


รูปที่ 4.30 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยใช้โครงข่ายนิรวลแบบไม่มีชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล้อง

ในการออกแบบโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนจะกำหนดให้จำนวนอินพุตของโครงข่ายเป็น 16 อินพุต ซึ่งได้ผลกระทบของจำนวนประเภทกีฬาที่เพิ่มขึ้นเทียบการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อนสำหรับข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.31 และ รูปที่ 4.32 ตามลำดับ



รูปที่ 4.31 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยใช้โครงข่ายนิรวลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล้องของข้อมูลฝึกฝน



รูปที่ 4.32 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับการเคลื่อนที่ของกล่องของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองข้อย่อย 0

จากผลการทดลองของการเคลื่อนที่ของกล่องเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นดังรูปที่ 4.28 และ รูปที่ 4.29 พบว่าไม่สามารถจำแนกประเภทกีฬาได้เนื่องจากลักษณะการเคลื่อนที่ของกล่องในกีฬาแต่ละประเภทมีความน่าจะเป็นในการเคลื่อนที่แต่ละทิศทางได้เท่ากันจำไม่สามารถจำแนกได้ ซึ่งเมื่อพิจารณาคอนฟิวดันเมตริกซ์ของปริภูมิส่วนประกอบमुखสำคัญเป็น 16 มิติและหน่วยนิเวศในชั้นซ่อนเป็น 28 หน่วยนิเวศ ดังตารางที่ 4.11 พบว่าให้โครงข่ายนิเวศให้คำตอบเป็นมวยไทยมากที่สุด ซึ่งอันจริงแล้วโครงข่ายนิเวศที่ออกแบบไว้ในโครงสร้างอื่นๆ จะให้คำตอบที่แตกต่างกันออกไป เช่น เมื่อจำนวนนิเวศในชั้นซ่อนเป็น 24 หน่วยนิเวศ โครงข่ายนิเวศจะให้คำตอบเป็น วอลเลย์บอลมากที่สุด เป็นต้น ซึ่งอธิบายได้ว่าโครงข่ายนิเวศพยายามที่จะคาดเดาประเภทกีฬาประเภทใดประเภทหนึ่งออกมา จึงเห็นได้จากผลการทดลองรูปที่ 4.31 และ รูปที่ 4.32 ความแม่นยำสำหรับ 2 ประเภทกีฬาเป็นประมาณร้อยละ 50 ความแม่นยำสำหรับ 3 ประเภทกีฬาเป็นประมาณร้อยละ 33 เป็นต้น

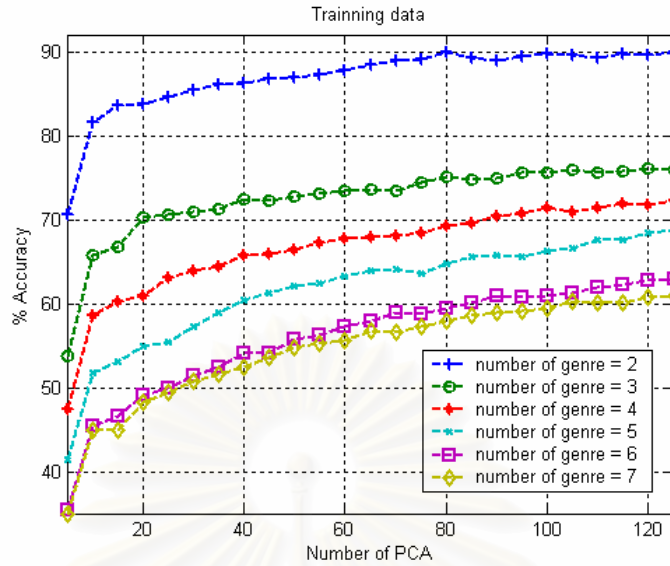
ตารางที่ 4.11 คอนฟิวชันเมทริกซ์ของการเคลื่อนที่ของกล้องสำหรับ 7 ประเภทกีฬา

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	34.53	40.84	0.00	0.20	11.51	8.40	4.50
มวยไทย	36.33	38.93	0.00	0.60	4.90	13.81	5.40
ฟุตบอล	36.63	45.74	0.10	0.30	5.70	6.60	4.90
กอล์ฟ	34.23	34.53	0.00	0.40	4.80	20.12	5.90
สปริงบอร์ด	36.23	45.24	0.00	0.70	1.00	10.91	5.90
เทนนิส	34.33	39.63	0.00	0.90	2.50	16.51	6.10
วอลเลย์บอล	34.63	54.15	0.10	1.00	0.40	6.10	3.60

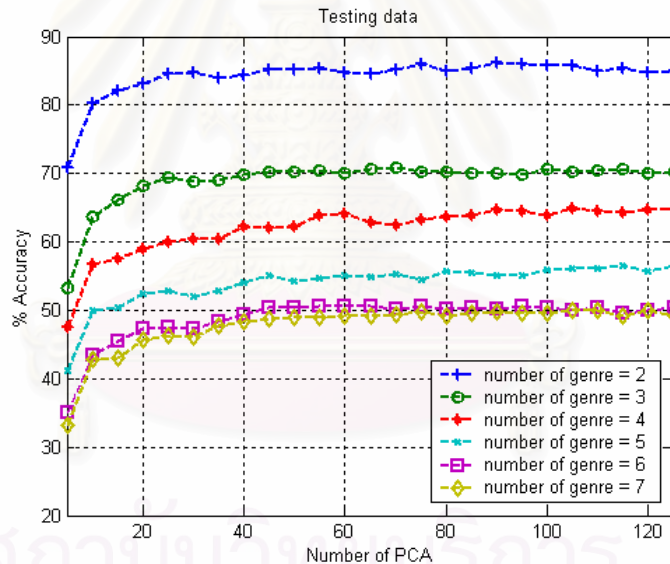
4.5.4 ผลการทดลองการจำแนกประเภทวิดีโอที่ศรัณกีฬาด้วยขอบสีโดยโครงข่ายนิรวลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ

ผลการทดลองลักษณะเฉพาะขอบสี ซึ่งภาพแต่ละภาพจะถูกลดขนาดจาก 384 x 448 จุดภาพ เป็น 96 x 112 จุดภาพและกำหนดขนาดบล็อกเป็น 16x16 จุดภาพ ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 126 มิติจำแนกด้วยการใช้โครงข่ายนิรวลที่ไม่มีชั้นซ่อนร่วมกับการใช้การวิเคราะห์ส่วนประกอบमुखสำคัญ เมื่อเพิ่มขนาดของปริภูมิส่วนประกอบमुखสำคัญได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.33 และ รูปที่ 4.34 ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย

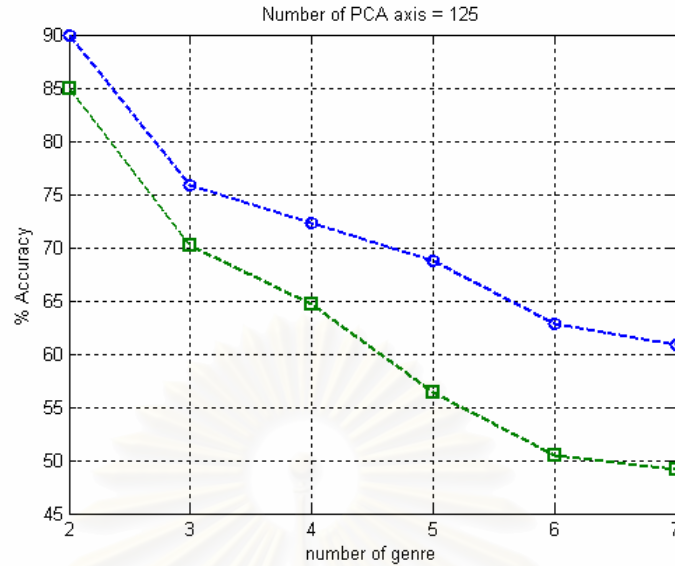


รูปที่ 4.33 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน



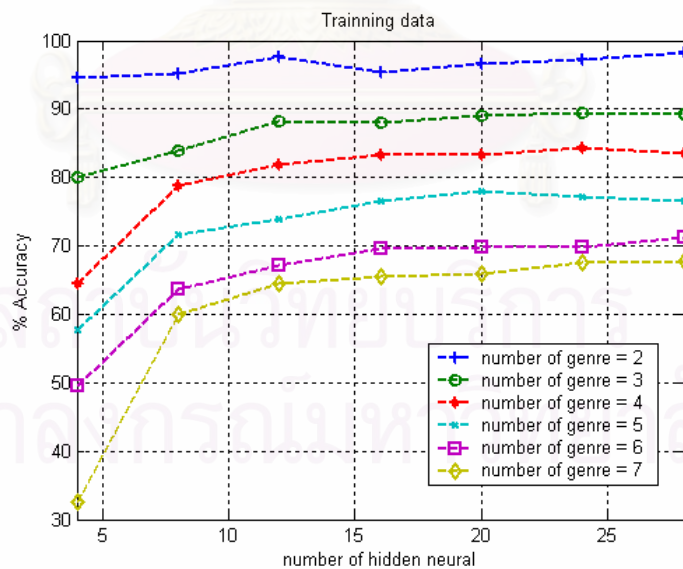
รูปที่ 4.34 ความแม่นยำของการจำแนกด้วยโครงข่ายนิรวลแบบไม่มีชั้นซ่อนร่วมกับการวิเคราะห์ ส่วนประกอบमुखสำคัญร่วม สำหรับ ขอบสีของข้อมูลทดสอบ

จากผลการทดลองดังรูปที่ 4.33 และ รูปที่ 4.34 ขนาดของปริภูมิที่เหมาะสมในการจำแนกหาได้จากขนาดปริภูมิที่ให้ค่าเฉลี่ยของความแม่นยำในการจำแนกทั้ง 7 ประเภทกีฬาที่สูงที่สุด สำหรับลักษณะเฉพาะขอบสี ได้ให้ผลความแม่นยำที่ดีที่สุดเมื่อปริภูมิมีขนาดเป็น 125 มิติ ซึ่งแสดงผลกระทบเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นได้ดังรูปที่ 4.35

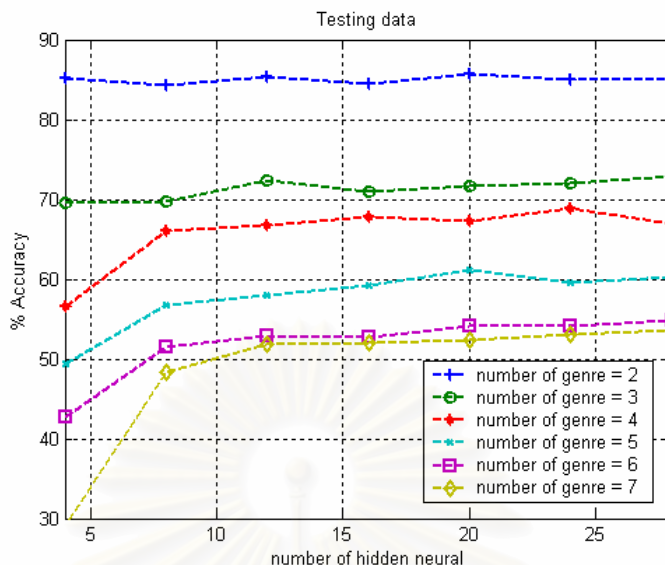


รูปที่ 4.35 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิรวลแบบไม่มีชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบที่สำคัญ สำหรับ ขอบสี

ในการออกแบบโครงข่ายนิรวลแบบมี 1 ชั้นซ่อนจะกำหนดให้จำนวนอินพุตของโครงข่ายเป็น 125 อินพุต ซึ่งได้ผลกระทบของจำนวนประเภทกีฬาที่เพิ่มขึ้นเทียบการเพิ่มจำนวนหน่วยนิรวลในชั้นซ่อน สำหรับข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.36 และ รูปที่ 4.37 ตามลำดับ



รูปที่ 4.36 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิรวลแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบที่สำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน



รูปที่ 4.37 ผลกระทบของการเพิ่มขึ้นประเภทกีฬาโดยการใช้โครงข่ายนิเวศแบบมี 1 ชั้นซ่อน ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ สำหรับ ขอบสีของข้อมูลฝึกฝน

วิเคราะห์และสรุปผลการทดลองหัวข้อย่อย 4.5.4

จากการทดลองลักษณะเฉพาะขอบสีจำแนกด้วยโครงข่ายนิเวศแบบไม่มีชั้นซ่อนได้ผลได้รูปที่ 4.33 และ รูปที่ 4.34 สำหรับข้อมูลฝึกฝนและข้อมูลทดสอบตามลำดับ จากผลการทดลองพบว่าเมื่อจำนวนประเภทกีฬาที่มีจำนวนมากขึ้น ขนาดปริภูมิของส่วนประกอบमुखสำคัญที่นำใช้ต้องมีมากขึ้นด้วย นอกจากนี้ผลความแม่นยำการจำแนกจะลดน้อยลงเมื่อจำนวนประเภทกีฬามีมากขึ้น ดังรูปที่ 4.35 เมื่อออกแบบโครงข่ายนิเวศให้ มีชั้นซ่อน 1 ชั้นซ่อน ได้ผลการทดลองดังรูปที่ 4.36 และ รูปที่ 4.37 สำหรับข้อมูลฝึกฝนและข้อมูลทดสอบตามลำดับ พบว่าการวิเคราะห์ส่วนประกอบमुखสำคัญสามารถปรับปรุงความแม่นยำเมื่อจำนวนประเภทกีฬาเพิ่มขึ้นได้ เมื่อใช้จำนวนหน่วยนิเวศที่มากเพียงพอ ซึ่งผลการทดลองที่ดีที่สุดของปริภูมิส่วนประกอบमुखสำคัญเป็น 125 มิติที่หน่วยนิเวศในชั้นซ่อนเป็น 28 หน่วยนิเวศ ซึ่งได้คอนฟิวชันเมทริกซ์ของขอบสีแสดงได้ดังตารางที่ 4.12

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.12 คอนฟิวชันเมทริกซ์ของขอบสีสำหรับ 7 ประเภทกีฬา

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	49.74	9.70	7.90	6.70	7.50	9.30	9.10
มวยไทย	6.80	66.86	7.60	6.40	2.70	3.50	6.10
ฟุตบอล	9.60	14.31	38.83	11.21	3.90	10.41	11.71
กอล์ฟ	6.70	3.10	7.10	66.66	4.80	9.90	1.70
สปริงบอร์ด	13.21	7.80	5.30	8.60	34.03	18.91	12.11
เทนนิส	11.61	4.10	6.40	10.61	9.30	52.85	5.10
วอลเลย์บอล	4.80	5.80	5.30	3.10	6.70	8.30	65.96

จากตารางที่ 4.12 คอนฟิวชันเมทริกซ์ พบว่าสามารถจำแนกกีฬาได้ดี 3 ประเภทกีฬาได้แก่มวยไทย กอล์ฟ และวอลเลย์บอลเนื่องจากมวยไทยมีลักษณะขอบสีจากขอบสนามที่เห็นอย่างชัดเจน ซึ่งแตกต่างกับกีฬาประเภทอื่นๆ ส่วน ฟุตบอลให้ผลการจำแนกที่น้อยที่สุดเนื่องจากกีฬาฟุตบอลให้ขอบสนามมีความคล้ายกับกีฬาประเภทอื่นเช่น เทนนิส หรือ บาสเก็ตบอลจึงสามารถจำแนกได้ผลไม่ดีนัก

สรุปการทดลอง

การทดลองในการทดลองที่ 5 พบว่าเมื่อเพิ่มประเภทกีฬาให้กับโครงข่ายนิวรอลร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญ จะทำผลความแม่นยำของการจำแนกประเภทกีฬาลดต่ำลงเนื่องจากความซับซ้อนของการจำแนกมีมากขึ้นทำให้ตัวจำแนกตัดสินใจได้ยากขึ้น ซึ่งจากการทดลองพบว่า ลักษณะเฉพาะให้ผลความแม่นยำของการจำแนกที่ดีที่สุด ซึ่งให้ผลดีในทุกประเภทกีฬา สำหรับฮิสโตแกรมสี ถ้านำมาใช้ในการจำแนกประเภทกีฬา ข้อมูลประเภทกีฬาที่นำมาให้ต้องมีความแตกต่างของสีสนามอย่างเห็นได้ชัดและจะต้องไม่รวมฉากเน้นตัวนักกีฬาและฉากผู้ชมเข้าไปในข้อมูลการฝึกฝนอีกด้วย สำหรับลักษณะเฉพาะของการเคลื่อนไหวที่ไม่สามารถ

นำมาจำแนกประเภทกีฬาได้เนื่องจากมีความน่าจะเป็นของการเคลื่อนที่ของกล้องในแต่ละประเภทกีฬามีความเป็นไปได้เท่ากัน ซึ่งโครงข่ายนิวรอลจะพยายามคาดเดากีฬาประเภทใดประเภทหนึ่งออกมา และสุดท้ายลักษณะเฉพาะของบาสเกตบอลที่สามารถจำแนกได้จำเป็นต้องมีขอบสีอย่างชัดเจน เช่นมวยไทย วอลเลย์บอล เทนนิส เป็นต้น

นอกจากนี้พบว่า การเพิ่มขึ้นของช่องให้กับโครงข่ายสามารถปรับปรุงประสิทธิภาพของการจำแนกประเภทกีฬาให้ถูกต้องเพิ่มมากขึ้นสำหรับทุกลักษณะเฉพาะอีกด้วย

4.6 การทดลองที่ 6 วิเคราะห์ความแม่นยำของการจำแนกประเภทวิดีโอที่บันทึกกีฬาด้วยตัวจำแนกปริภูมิคู่ศูนย์

เพื่อปรับปรุงของความแม่นยำของการจำแนกให้มีประสิทธิภาพ วิทยานิพนธ์นี้ได้นำเสนอเทคนิคอีกเทคนิคหนึ่งซึ่งเรียกว่า ตัวจำแนกโดยใช้ปริภูมิคู่ศูนย์ เนื่องจากปริภูมิคู่ศูนย์เป็นปริภูมิที่เกิดจากเมทริกซ์การแปลงที่ทำให้การกระจายตัวระหว่างกลุ่มมีระยะห่างกันมากที่สุด ในขณะที่การกระจายตัวภายในกลุ่มชนิดเดียวกันเข้าใกล้กันมากที่สุด นั่นหมายความว่าสามารถจำแนกข้อมูลได้ง่ายขึ้นเมื่อสร้างตัวจำแนกในปริภูมิคู่ศูนย์ ดังนั้นในการทดลองนี้จะศึกษาการจำแนกโดยใช้ตัวจำแนกปริภูมิคู่ศูนย์

วัตถุประสงค์

เพื่อทดสอบประสิทธิภาพของตัวจำแนกในปริภูมิคู่ศูนย์ ด้วยการใช้อัลกอริทึมการฝึกสอนแบบไม่มีผู้สอน การเคลื่อนที่ของกล้อง และ ขอบสี เมื่อจำนวนประเภทกีฬามีจำนวนมากขึ้น

ปัจจัยกำหนดในการทดลอง

ในการทดลองที่ 6 นี้เป็นการศึกษาถึงผลกระทบเมื่อจำนวนประเภทกีฬามีจำนวนมากขึ้นใน 4 ลักษณะเฉพาะ เมื่อจำแนกด้วยปริภูมิคู่ศูนย์ ซึ่งในการทดลองนี้แบ่งเป็นหัวข้อย่อย 4 ส่วนตามลักษณะเฉพาะที่ได้นำเสนอในวิทยานิพนธ์ในบทที่ 3 มีปัจจัยในการทดลองคล้ายกับการทดลองที่ 5 ดังต่อไปนี้

- ประเภทกีฬาสำหรับทดลองเป็น 7 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด เทนนิส และ วอลเลย์บอล ที่ประกอบไปด้วย ฉากการแข่งขัน ฉากเน้นไปตวันนักกีฬา และ ฉากผู้ชมจำนวน 27 วิดิทัศน์เป็นเวลาประมาณ 233 นาที โดยในหนึ่งข้อดประกอบด้วย 50 เฟรม ซึ่งจำนวนข้อมูลในการฝึกฝนและทดสอบแสดงได้ดังตารางที่ 4.8

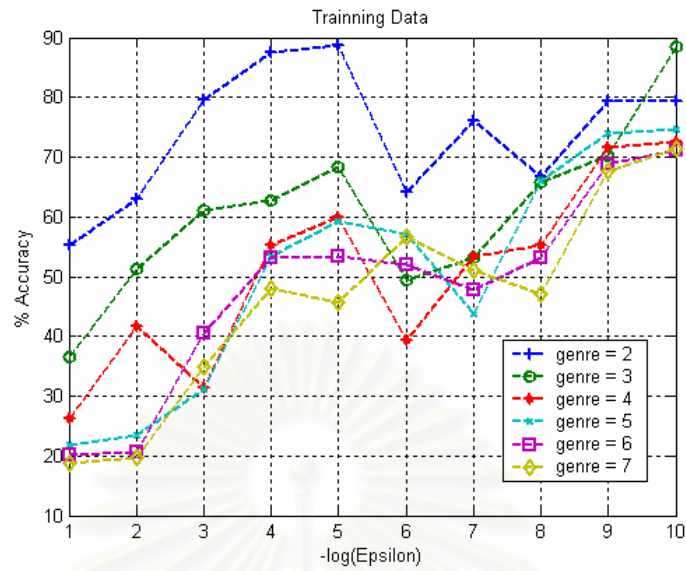
- จำนวนประเภทกีฬาที่เพิ่มขึ้นมีลำดับดังต่อไปนี้
 - จำนวน 2 ประเภทกีฬา ได้แก่ บาสเก็ตบอล และ มวยไทย
 - จำนวน 3 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย และ ฟุตบอล
 - จำนวน 4 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล และ กอล์ฟ
 - จำนวน 5 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ และ สปริงบอร์ด
 - จำนวน 6 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด และ เทนนิส
 - จำนวน 7 ประเภทกีฬา ได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ สปริงบอร์ด เทนนิส และ วอลเลย์บอล
- อัตราส่วนระหว่างจำนวนข้อมูลเรียนรู้ต่อจำนวนข้อมูลวาไลเดต (Validate data) เป็น 4 ต่อ 1 ซึ่งหยุดการฝึกฝนเมื่อค่าความผิดพลาดที่ได้จากข้อมูลวาไลเดต (Validate data) มีแนวโน้มสูงขึ้น

ผลการทดลอง

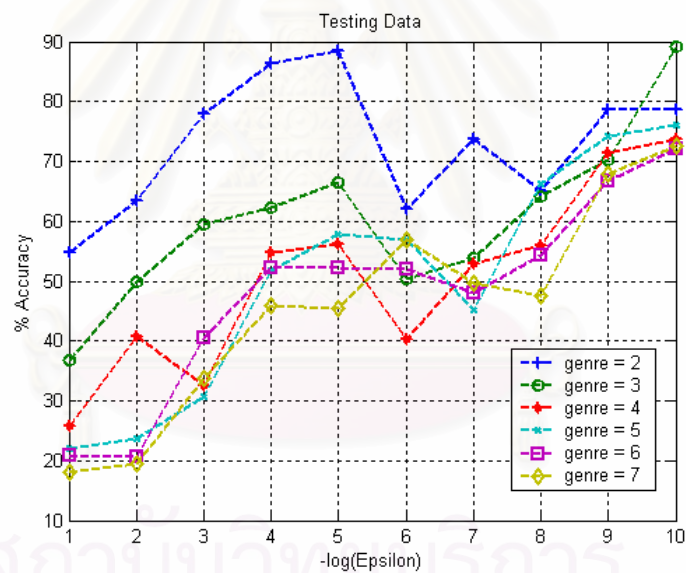
ซึ่งในการทดลองนี้แบ่งได้ผลการทดลอง 4 ผลการทดลองย่อยตามลักษณะเฉพาะ ดังนี้

4.6.1 ผลการทดลองการจำแนกประเภทวิถีทัศน์กีฬาด้วยการใช้ฮิสโตแกรมสีโดยตัวจำแนกในปริภูมิสีคู่ขนาน

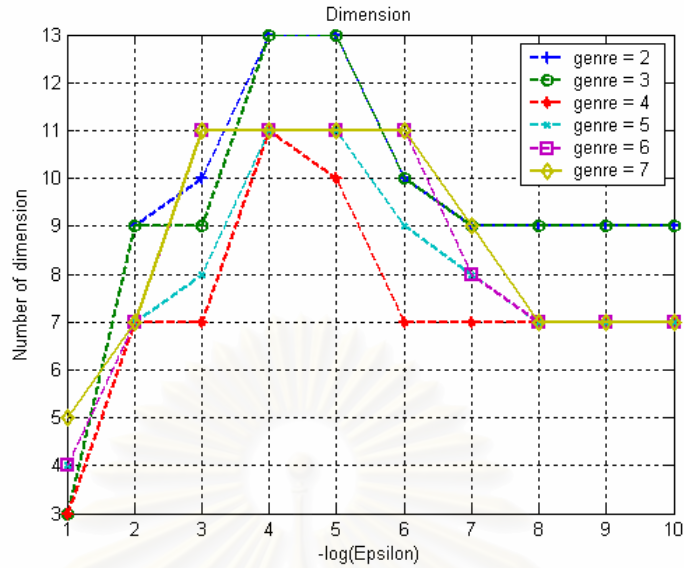
ผลการทดลองลักษณะเฉพาะฮิสโตแกรมขนาด 64 ถังสี จำแนกด้วยการใช้ตัวจำแนกในปริภูมิสีคู่ขนานเมื่อเลือกขนาดปริภูมิสีคู่ขนานโดยการปรับ ϵ ได้ผลการทดลองสำหรับข้อมูลฝึกฝนและทดสอบดังรูปที่ 4.38 และ รูปที่ 4.39 ตามลำดับ และ ขนาดปริภูมิสำหรับฮิสโตแกรมสีที่นำมาใช้เมื่อปรับ ϵ ดังรูปที่ 4.40



รูปที่ 4.38 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ฮิสโตแกรมสี่จากข้อมูลฝึกฝน

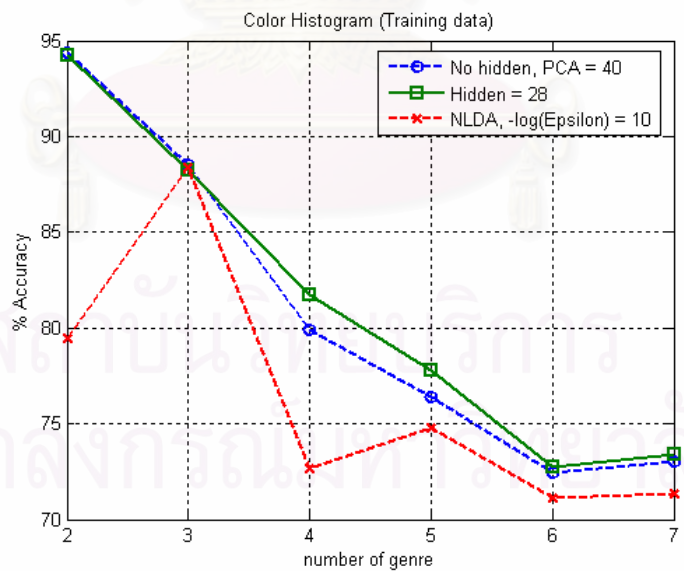


รูปที่ 4.39 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ฮิสโตแกรมสี่จากข้อมูลทดสอบ

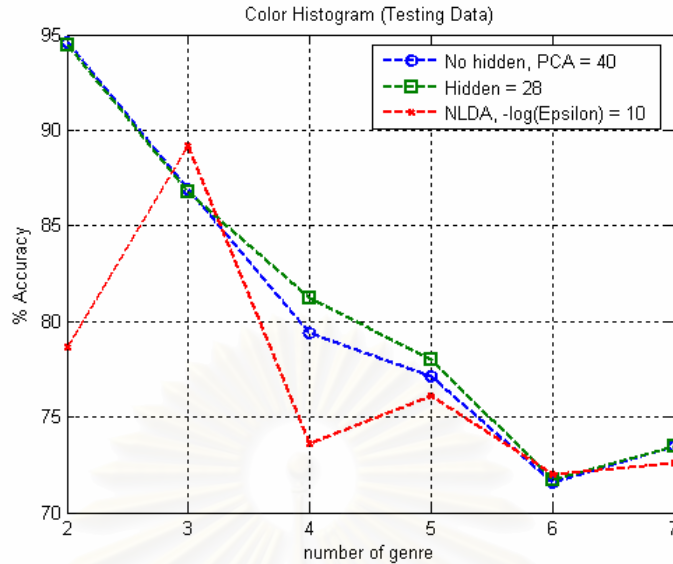


รูปที่ 4.40 ขนาดปริภูมิสุญญ์ของฮิสโตแกรมสี

เมื่อเปรียบเทียบกับ การทดลองที่ 6 โดยใช้การวิเคราะห์ส่วนประกอบหลัก ร่วมกับโครงข่ายนิรวลในกรณีที่ไม่มีการซ่อน และ มีการซ่อนของข้อมูลฝึกฝนและทดสอบได้รูปที่ 4.41 และ รูปที่ 4.42 ตามลำดับ



รูปที่ 4.41 การเปรียบเทียบตัวจำแนก สำหรับ ฮิสโตแกรมสีของข้อมูลฝึกฝน



รูปที่ 4.42 การเปรียบเทียบตัวจำแนก สำหรับ ฮิสโตแกรมสีของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองห่วยย่อย 4.6.1

จากการทดลองดังรูปที่ 4.38 และ รูปที่ 4.39 พบว่าเมื่อปรับค่า ϵ มีแนวโน้มของความแม่นยำสูงขึ้นเนื่องจากขนาดของปริภูมิในการทำการแปลงหาปริภูมิหลักมีจำนวนมากขึ้นดังรูปที่ 4.40 จึงทำให้ความแม่นยำสูงขึ้น แต่เมื่อค่า ϵ มีค่าน้อยกว่า 10^{-5} การฉายในปริภูมิสูงทำให้ขนาดของปริภูมิมีขนาดน้อยลงจึงทำให้ผลการจำแนกน้อยลง ซึ่งจากเห็นได้จากรูปที่ 4.40 มีขนาดปริภูมิที่ลดลงจึงทำให้ผลความแม่นยำของการจำแนกที่น้อยลง ซึ่งจากรูปที่ 4.47 เป็นการเปรียบเทียบตัวจำแนกในปริภูมิสูงสุญญ์ กับ โครงข่ายนิเวรอลพบว่า โครงข่ายนิเวรอลให้ผลความแม่นยำการจำแนกที่ดีกว่า แต่เมื่อพิจารณาจำนวนพารามิเตอร์ในการใช้และความรวดเร็วฝึกฝนแล้วตัวจำแนกในปริภูมิสูงสุญญ์จะให้ประสิทธิภาพที่ดีกว่า เนื่องจากใช้เวลาในการฝึกฝนที่น้อยกว่าและพารามิเตอร์ที่เก็บในหน่วยความจำที่น้อยกว่า แต่ให้ผลความแม่นยำที่ใกล้เคียงกับโครงข่ายนิเวรอล ซึ่งคอนฟิวชันเมทริกซ์ของปริภูมิสูงสุญญ์สำหรับฮิสโตแกรม แสดงได้ดังตารางที่ 4.13 ซึ่งจะให้คำตอบเป็นก็ฟากอล์ฟมากที่สุด

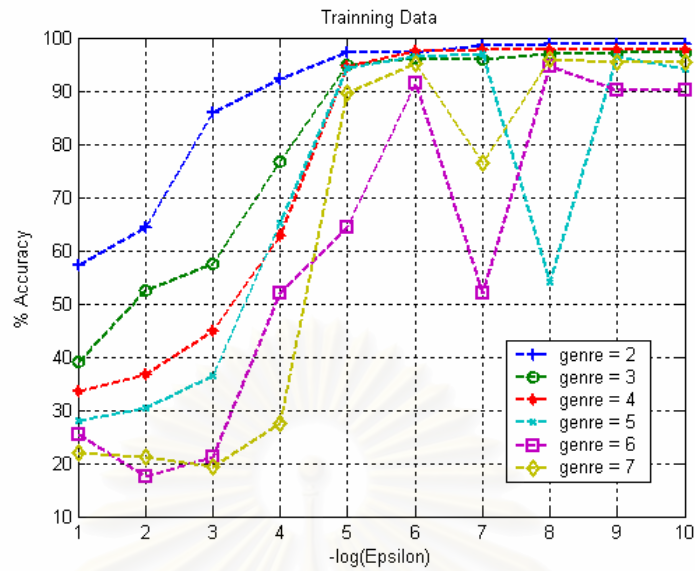
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.13 คอนฟิวชันเมทริกซ์ของปริภูมิคู่ศูนย์ สำหรับฮิสโตแกรม 7 ประเภทกีฬา

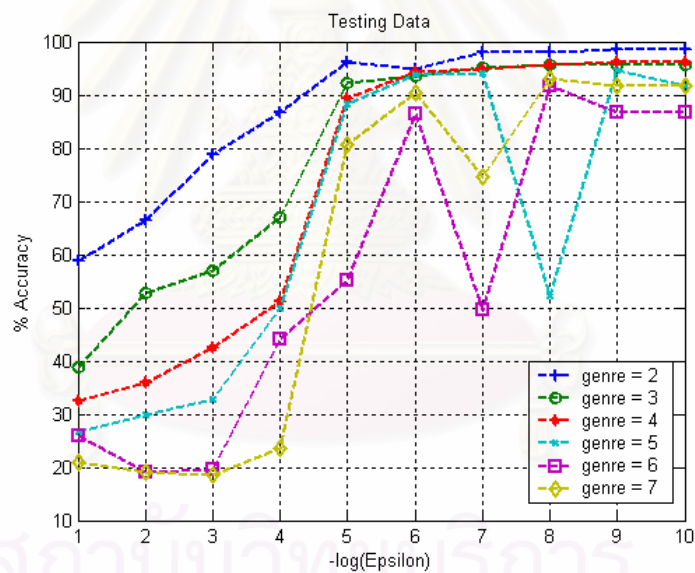
	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	75.07	0.00	3.60	3.90	6.30	11.11	0.00
มวยไทย	4.50	73.87	4.20	3.00	1.80	12.61	0.00
ฟุตบอล	7.20	0.90	53.15	27.62	5.70	4.50	0.90
กอล์ฟ	1.20	0.30	4.50	87.08	2.10	4.80	0.00
สปริงบอร์ด	1.50	0.00	3.30	7.20	75.67	11.11	1.20
เทนนิส	2.70	0.60	0.90	4.20	14.41	76.87	0.30
วอลเลย์บอล	0.30	0.00	0.60	7.50	19.51	5.70	66.36

4.6.2 ผลการทดลองการจำแนกประเภทวิถีทัศนกีฬาดำเนินการใช้ออโตคอร์รีโลแกรมโดยตัวจำแนกในปริภูมิคู่ศูนย์

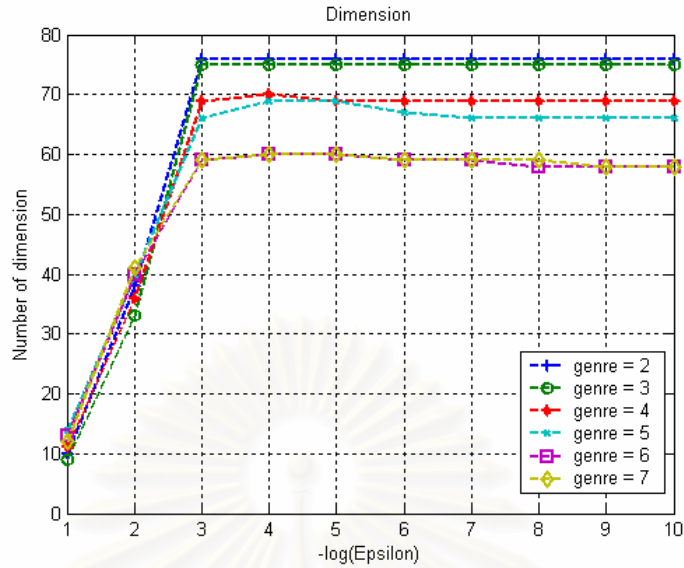
ผลการทดลองลักษณะเฉพาะออโตคอร์รีโลแกรม ซึ่งภาพแต่ละภาพจะถูกลดขนาดจาก 384 x 448 จุดภาพ เป็น 96 x 112 จุดภาพและกำหนดเซตของระยะห่างเป็น $k = \{1,3,5,7,9\}$ ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 320 มิติจำแนกด้วยการใช้ปริภูมิคู่ศูนย์ เมื่อเลือกขนาดปริภูมิคู่ศูนย์โดยการปรับ ε ได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.43 และ รูปที่ 4.44 ตามลำดับ และ ขนาดปริภูมิสำหรับออโตคอร์รีโลแกรมที่นำมาใช้เมื่อปรับ ε ดังรูปที่ 4.45



รูปที่ 4.43 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ออโตคอเรียไลแกรมจากข้อมูลฝึกฝน

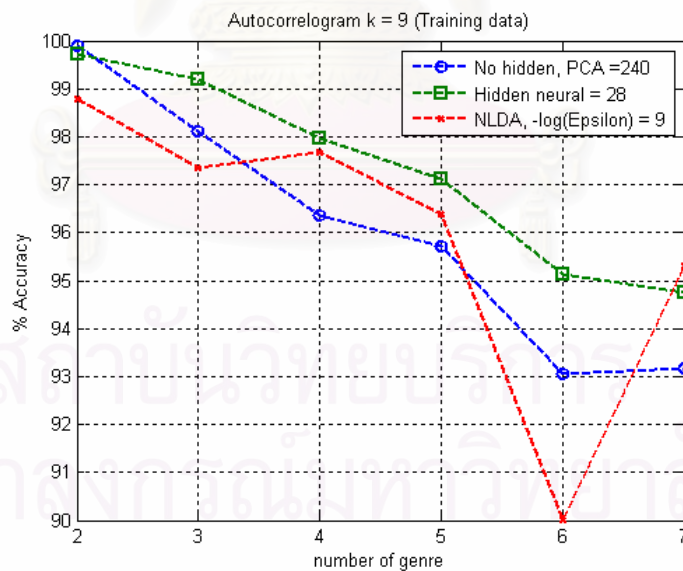


รูปที่ 4.44 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ออโตคอเรียไลแกรมจากข้อมูลทดสอบ

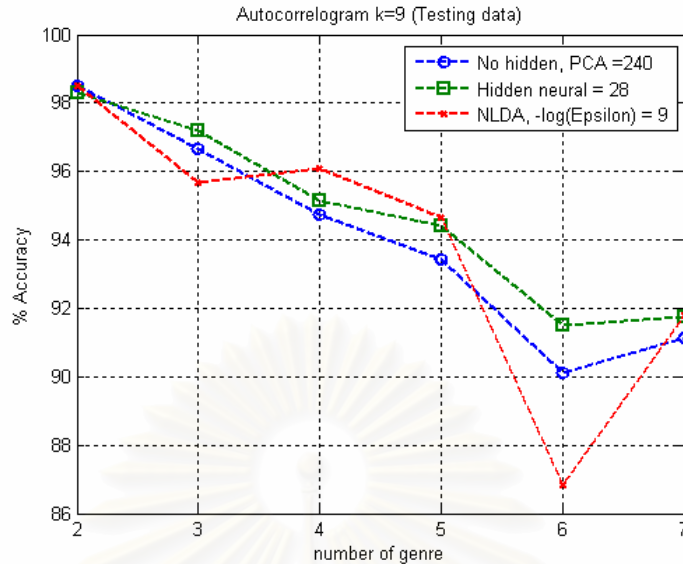


รูปที่ 4.45 ขนาดปริภูมิคู่ศูนย์ของออโตคอรัลแกรม

เมื่อเปรียบเทียบกับ การทดลองที่ 6 โดยใช้การวิเคราะห์ส่วนประกอบหลัก ร่วมกับโครงข่ายนิเวศในกรณีที่ไม่มีชั้นซ่อน และมีชั้นซ่อนของข้อมูลฝึกฝนและทดสอบได้รูปที่ 4.46 และ รูปที่ 4.7 ตามลำดับ



รูปที่ 4.46 การเปรียบเทียบตัวจำแนก สำหรับ ออโตคอรัลแกรมของข้อมูลฝึกฝน



รูปที่ 4.47 การเปรียบเทียบตัวจำแนก สำหรับ ออโตคอร์รีโลแกรมของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองหน่วยย่อย 4.6.2

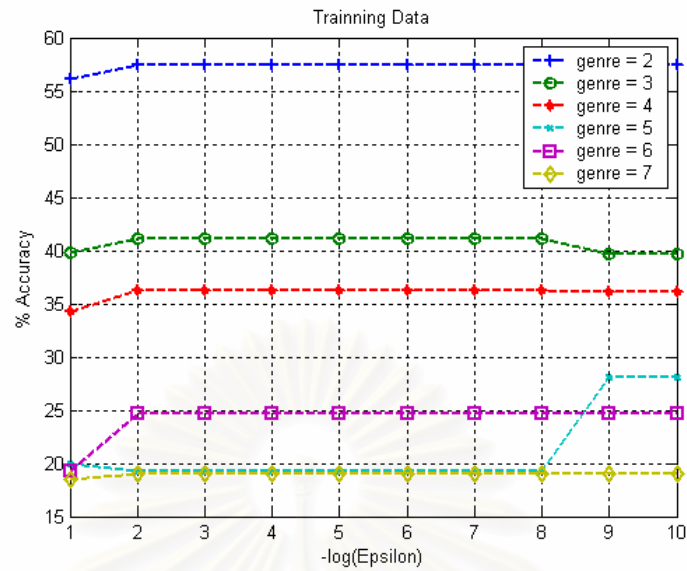
จากการทดลองดังรูปที่ 4.43 และ รูปที่ 4.44 พบว่าเมื่อปรับค่า ϵ มีแนวโน้มของความแม่นยำสูงขึ้น เนื่องจากขนาดของปริภูมิในการทำการแปลงหาปริภูมิหลักมีจำนวนมากขึ้น ดังรูปที่ 4.45 จึงทำให้ความแม่นยำสูงขึ้น แต่เมื่อค่า ϵ มีค่าน้อยกว่า 10^{-5} จะมีแนวโน้มคงที่ ซึ่งจากเห็นได้จากรูปที่ 4.45 มีขนาดปริภูมิที่ใช้มีขนาดที่คงที่ ซึ่งจากรูปที่ 4.47 เป็นการเปรียบเทียบตัวจำแนกในปริภูมิสูงสุญญ์ กับ โครงข่ายนิเวรอลพบว่า โครงข่ายนิเวรอลให้ผลความแม่นยำการจำแนกที่ดีกว่า แต่เมื่อพิจารณาจำนวนพารามิเตอร์ในการใช้และความรวดเร็วในการฝึกฝนแล้วตัวจำแนกในปริภูมิสูงสุญญ์จะให้ประสิทธิภาพที่ดีกว่า เนื่องจากใช้เวลาน้อยกว่าและพารามิเตอร์ที่เก็บในหน่วยความจำที่น้อยกว่าและให้ผลความแม่นยำที่ใกล้เคียงกับโครงข่ายนิเวรอล อีกทั้งขนาดปริภูมิที่มีขนาดเล็กกว่า ซึ่งจากตารางที่ 4.14 พบว่าออโตคอร์รีโรแกรมสามารถให้ความแม่นยำของการจำแนกได้สำหรับทุกประเภทกีฬา

ตารางที่ 4.14 คอนฟิวชันเมทริกซ์ของปริภูมิสี่ศูนย์ สำหรับบอโตคอรรีโโลแกรม 7 ประเภทกีฬา

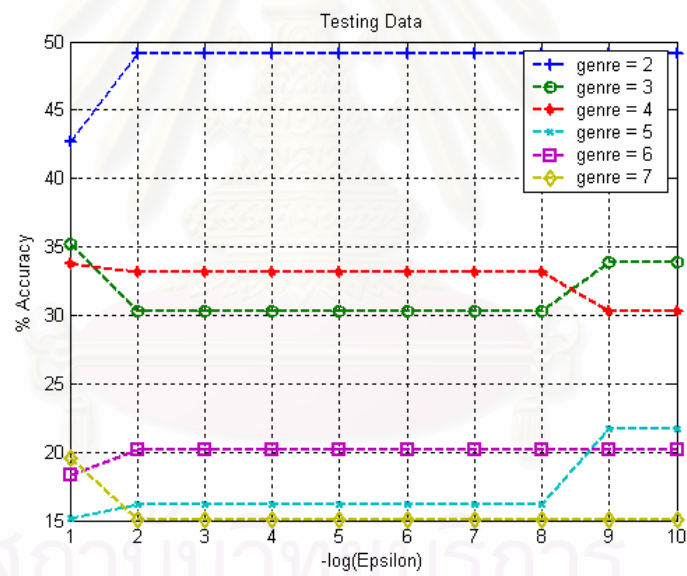
	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	91.89	0.90	1.80	0.00	2.70	2.70	0.00
มวยไทย	0.90	95.49	0.30	0.60	0.60	1.80	0.30
ฟุตบอล	2.10	1.80	86.18	1.20	5.40	3.00	0.30
กอล์ฟ	0.00	0.90	0.90	92.79	4.80	0.60	0.00
สปริงบอร์ด	0.00	0.00	0.60	1.50	93.99	3.90	0.00
เทนนิส	1.50	1.580	0.90	0.00	5.10	90.69	0.30
วอลเลย์บอล	0.90	0.60	0.60	0.00	0.90	5.40	91.59

4.6.3 ผลการทดลองการจำแนกประเภทวิถีทัศนกีฬาดำเนินการด้วยการเคลื่อนที่ของกล่องโดยตัวจำแนกในปริภูมิสี่ศูนย์

ผลการทดลองลักษณะเฉพาะการเคลื่อนที่ของกล่อง ซึ่งตั้งลักษณะการเคลื่อนที่ของกล่องทุก 5 เฟรม ใน 1 ซ็อต ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 20 มิติจำแนกด้วยการใช้ตัวจำแนกในปริภูมิสี่ศูนย์เมื่อเลือกขนาดปริภูมิสี่ศูนย์โดยการปรับ ϵ ได้ผลการทดลองสำหรับข้อมูลฝึกฝนและทดสอบรูปที่ 4.48 และ รูปที่ 4.9 ตามลำดับ และ ขนาดปริภูมิสำหรับลักษณะเฉพาะของการเคลื่อนที่ของกล่องที่นำมาใช้เมื่อปรับ ϵ ดังรูปที่ 4.50

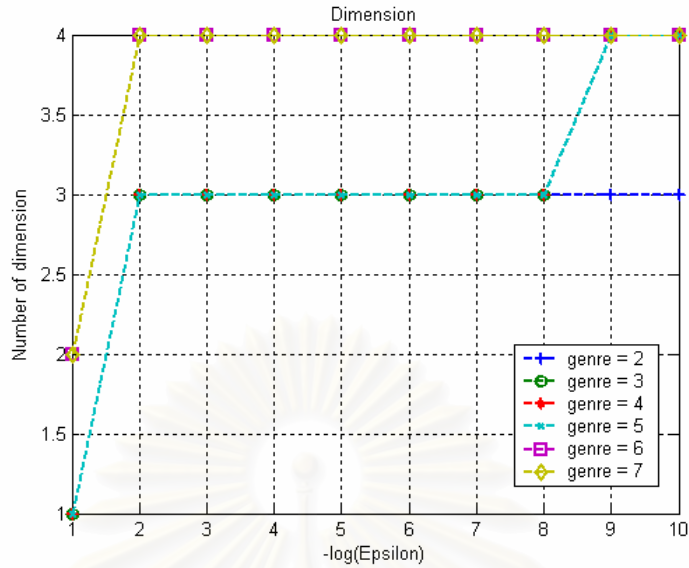


รูปที่ 4.48 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ การเคลื่อนที่ของกล่องจากข้อมูลฝึกฝน



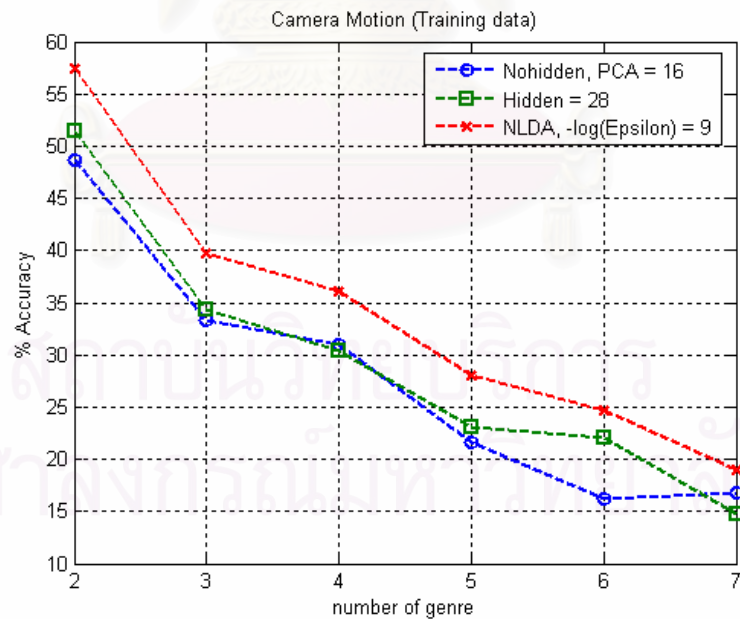
รูปที่ 4.49 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ การเคลื่อนที่ของกล่องจากข้อมูลทดสอบ

สถาบันนวัตกรรม
จุฬาลงกรณ์มหาวิทยาลัย

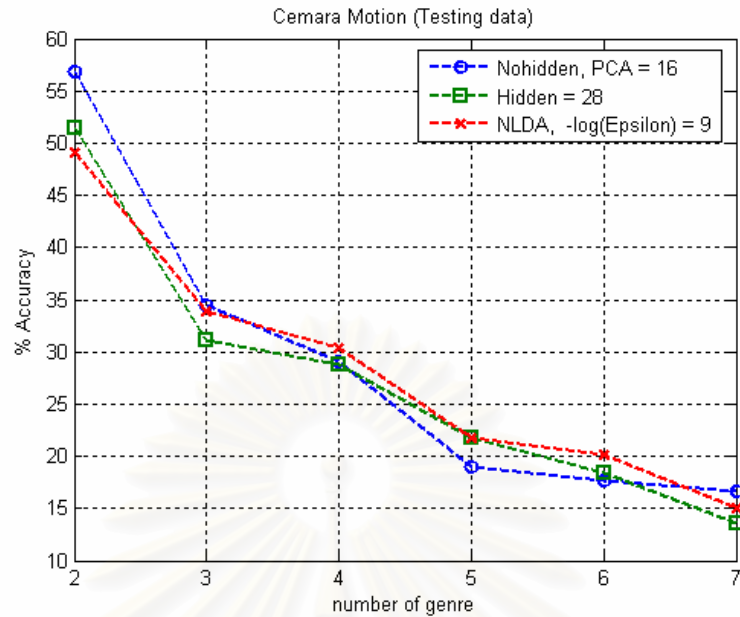


รูปที่ 4.50 ขนาดปริภูมิผู้ศูนย์ของการเคลื่อนที่ของกล้อง

เมื่อเปรียบเทียบกับ การทดลองที่ 6 โดยใช้การวิเคราะห์ส่วนประกอบที่สำคัญ ร่วมกับโครงข่ายนิรอรอลในกรณีที่ไม่มีการซ่อน และ มีการซ่อนของข้อมูลฝึกฝนและทดสอบได้รูปที่ 4.51 และ รูปที่ 4.52 ตามลำดับ



รูปที่ 4.51 การเปรียบเทียบตัวจำแนกโดยใช้การเคลื่อนที่ของกล้องของข้อมูลฝึกฝน



รูปที่ 4.52 การเปรียบเทียบตัวจำแนกโดยใช้การเคลื่อนที่ของกล้องของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองหน่วยย่อย 0

ผลการทดลองลักษณะเฉพาะการเคลื่อนที่ของกล้องรูปที่ 4.48 และ รูปที่ 4.49 พบว่าแนวโน้มของความแม่นยำคงที่ ซึ่งเป็นการคาดเดาคำตอบของตัวจำแนกในปริภูมิสูงสู้นั้น เมื่อพิจารณาขนาดของปริภูมิสูงสู้นั้นดัง รูปที่ 4.50 พบว่า ขนาดของปริภูมิที่นำไปใช้เป็น 3 ถึง 4 มิติ ซึ่งพบเพียงการฉายลงปริภูมิหลักเท่านั้น ไม่มีการฉายลงในปริภูมิสูงสู้นั้น เนื่องจากลักษณะเฉพาะการเคลื่อนที่ของกล้องให้ค่าเวกเตอร์ลักษณะ (eigen-value) ที่มาก ทำให้เกิดปริภูมิสูงสู้นั้น ดังนั้น ปริภูมิสูงสู้นั้นไม่สามารถจำแนกประเภทกีฬาได้ สำหรับลักษณะการเคลื่อนที่ของกล้อง ซึ่งแสดงคอนฟิวชันเมทริกซ์ได้ดังตารางที่ 4.15

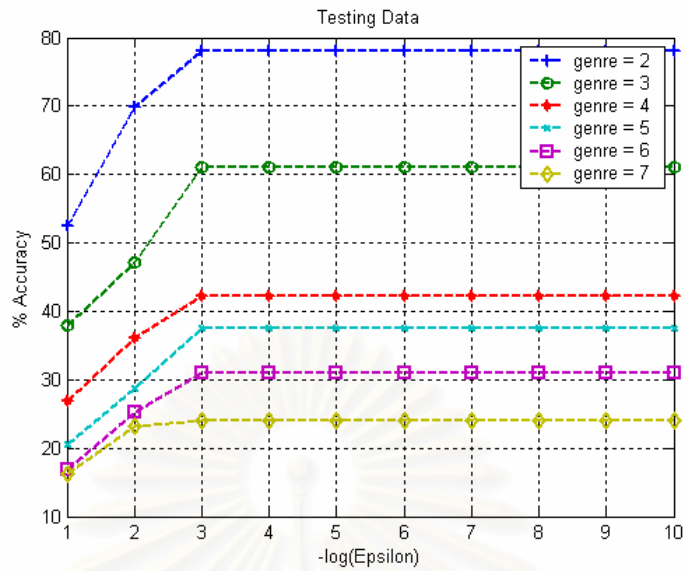
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.15 คอนฟิวชันเมทริกซ์ของปริภูมิคู่ศูนย์ สำหรับการเคลื่อนที่กลอง 7 ประเภทกีฬา

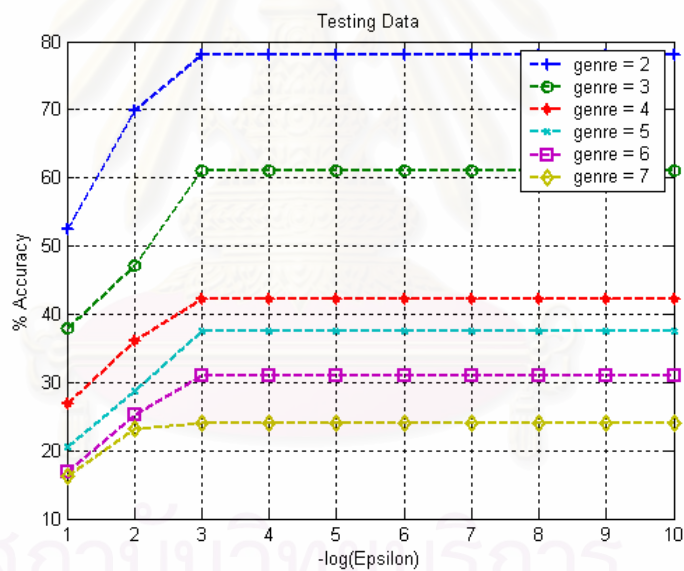
	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	15.31	26.72	12.91	34.83	0.90	6.30	3.00
มวยไทย	17.41	18.61	24.32	14.71	9.90	1.20	13.81
ฟุตบอล	20.10	21.92	19.21	18.31	4.20	11.71	4.50
กอล์ฟ	6.30	15.31	17.11	49.84	6.00	2.10	3.30
สปริงบอร์ด	21.32	21.92	25.82	24.02	0.00	6.90	0.00
เทนนิส	16.51	16.81	17.71	35.13	5.70	2.40	5.70
วอลเลย์บอล	28.52	27.92	20.42	13.21	0.00	9.90	0.00

4.6.4 ผลการทดลองการจำแนกประเภทวิถีทัศน์กีฬาด้วยการใช้ขอบสีโดยตัวจำแนกในปริภูมิคู่ศูนย์

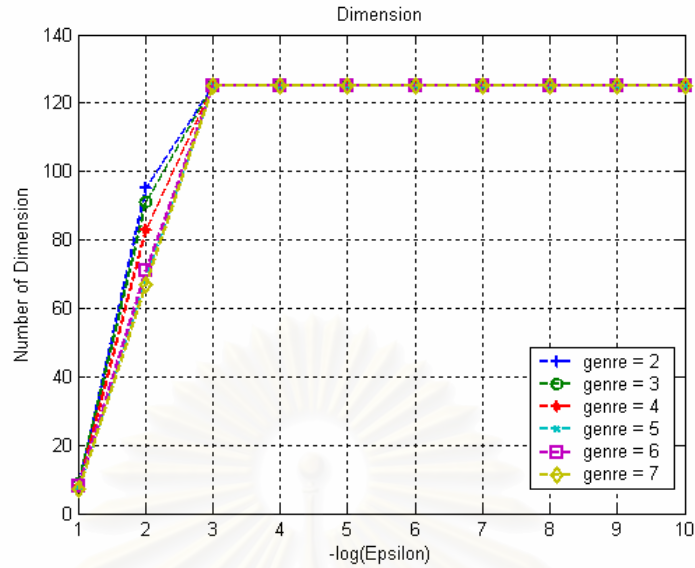
ผลการทดลองลักษณะเฉพาะฮอโตคอรีโโลแกรม ซึ่งภาพแต่ละภาพจะถูกลดขนาดจาก 384 x 448 จุดภาพ เป็น 96 x 112 จุดภาพและกำหนดขนาดบล็อกเป็น 16x16 จุดภาพ ดังนั้นขนาดของเวกเตอร์ลักษณะเฉพาะเป็น 126 มิติจำแนกด้วยการใช้ปริภูมิคู่ศูนย์เมื่อเลือกขนาดปริภูมิคู่ศูนย์โดยการปรับ ε ได้ผลการทดลองของข้อมูลฝึกฝนและข้อมูลทดสอบดังรูปที่ 4.53 และ รูปที่ 4.54 และ ขนาดปริภูมิสำหรับลักษณะเฉพาะขอบสีที่นำมาใช้เมื่อปรับ ε ดังรูปที่ 4.55



รูปที่ 4.53 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ขอบสีจากข้อมูลฝึกฝน

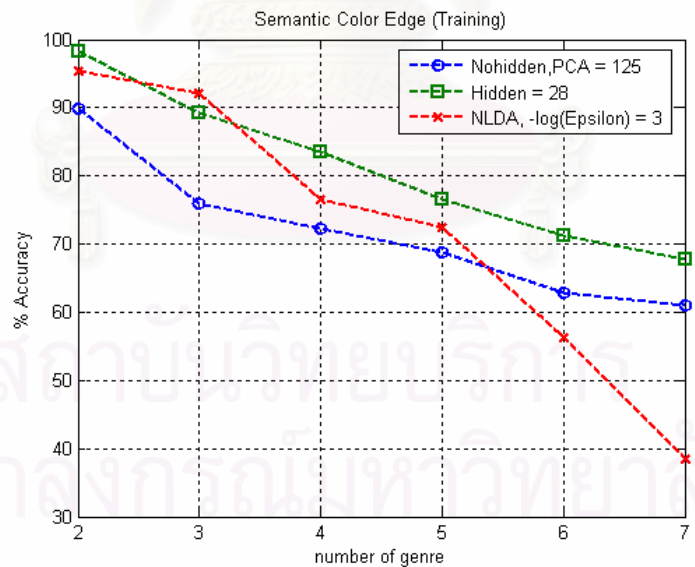


รูปที่ 4.54 ผลการทดลองตัวจำแนกปริภูมิคู่ศูนย์ สำหรับ ขอบสีจากข้อมูลทดสอบ

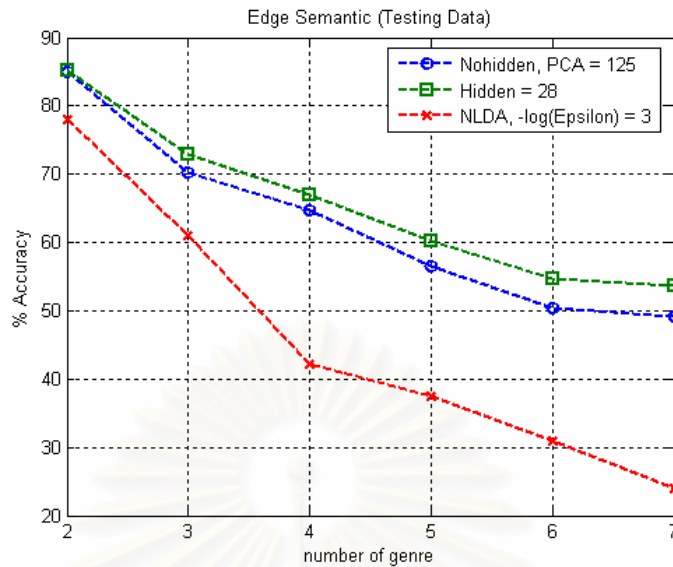


รูปที่ 4.55 ขนาดปริภูมิคู่ศูนย์ของขอบสี

เมื่อเปรียบเทียบกับ การทดลองที่ 6 โดยใช้การวิเคราะห์ส่วนประกอบที่สำคัญ ร่วมกับโครงข่ายนิเวศในกรณีที่ไม่มีชั้นซ่อน และมีชั้นซ่อนของข้อมูลฝึกฝนและทดสอบได้รูปที่ 4.56 และ รูปที่ 4.57 ตามลำดับ



รูปที่ 4.56 การเปรียบเทียบตัวจำแนก สำหรับ ขอบสีของข้อมูลฝึกฝน



รูปที่ 4.57 การเปรียบเทียบตัวจำแนก สำหรับ ขอบสีของข้อมูลทดสอบ

วิเคราะห์และสรุปผลการทดลองห่วยย่อย 0

จากการทดลองดังรูปที่ 4.53 และรูปที่ 4.54 พบว่าเมื่อปรับให้ ϵ มีน่าน้อยลงจะทำให้ผลความแม่นยำของการจำแนกเพิ่ม สำหรับทุกการเพิ่มของประเภทกีฬาเนื่องจากขนาดของปริภูมิหลักมีจำนวนมากขึ้น และมีแนวโน้มคั้งที่เมื่อ $\epsilon < 10^{-3}$ เนื่องจากขนาดของปริภูมิดังรูปที่ 4.55 มีแนวโน้มคั้งที่ ซึ่งเกิดจากไม่มีปริภูมิคู่ศูนย์สำหรับลักษณะเฉพาะขอบสี สาเหตุจากเมทริกซ์การกระจายภายในกลุ่มมีการกระจายตัวของข้อมูลสูงจึงทำให้ค่าลักษณะเฉพาะที่เกิดจากเมทริกซ์การกระจายในกลุ่มมีค่าสูง ดังนั้นจึงไม่ทำให้ขนาดของปริภูมิลดลงเมื่อจำนวนประเภทกีฬาเพิ่มมากขึ้น นอกจากนี้พบว่าเมื่อจำนวนประเภทกีฬามีมากขึ้น ผลความแม่นยำของการจำแนกจะต่ำลงดังรูปที่ 4.56 และ รูปที่ 4.57 ซึ่งตัวจำแนกในปริภูมิคู่ศูนย์ให้ผลที่แยกว่าโครงข่ายนิรอรลทั้งแบบมีไม่มีชั้นซ่อนและแบบมี 1 ชั้นซ่อน สามารถแสดงคอนฟิวชันเมทริกซ์ของปริภูมิคู่ศูนย์ สำหรับขอบสีได้ดังตารางที่ 4.16 พบว่า ขอบสีสามารถจำแนกประเภทกีฬาที่แสดงด้วยลักษณะของขอบสนามอย่างชัดเจน เช่น บาสเก็ตบอล เทนนิส หรือ มวยไทย ตามลำดับ แต่ไม่สามารถจำแนกกีฬา กอล์ฟ และ สปริงบอร์ดได้เนื่องจากไม่มีขอบของสนามที่เห็นอย่างชัดเจน

ตารางที่ 4.16 คอนฟิวชันเมทริกซ์ของปริมาณผู้ศูนย์ สำหรับขอบสี 7 ประเภทกีฬา

	ร้อยละความแม่นยำที่จำแนกได้ในแต่ละประเภทกีฬา						
	บาสเก็ตบอล	มวยไทย	ฟุตบอล	กอล์ฟ	สปริงบอร์ด	เทนนิส	วอลเลย์บอล
บาสเก็ตบอล	48.04	12.91	3.90	2.70	0.00	31.53	0.90
มวยไทย	8.10	37.83	2.40	15.01	0.00	36.33	0.30
ฟุตบอล	17.71	10.51	14.11	9.90	0.00	46.24	1.50
กอล์ฟ	17.41	8.70	14.71	7.50	0.00	51.65	0.00
สปริงบอร์ด	23.42	13.81	3.60	6.60	1.20	50.75	0.60
เทนนิส	3.63	8.70	7.50	0.60	0.00	46.54	0.00
วอลเลย์บอล	12.91	11.41	2.70	13.51	0.00	47.14	12.31

4.7 สรุปผลการทดลอง

ในบทนี้ได้กล่าวถึงการทดลองการจำแนกประเภทวิถีทัศนกีฬาโดยใช้ลักษณะเฉพาะระดับล่างแบบต่างๆ กับตัวจำแนกที่นำเสนอ โดยการทดลองที่ 1 ถึงการทดลองที่ 4 เป็นการทดลองเพื่อวิเคราะห์ประสิทธิภาพของการจำแนกจากลักษณะเฉพาะระดับล่างของวิถีทัศน กับตัวจำแนกที่นำเสนอ จากการทดลองพบว่า ลักษณะเฉพาะสี่จะให้ผลความแม่นยำที่ดีว่าลักษณะเฉพาะระดับล่างอื่นๆ โดยที่ออกโตคอร์รีโลแกรมให้ผลที่ดีที่สุดสำหรับ 4 ประเภทกีฬาได้แก่ มวยไทย ฟุตบอล เทนนิส และ วอลเลย์บอล ซึ่งออกโตคอร์รีโลแกรมสามารถแทนจากเน้นักกีฬา และ จากผู้ชม ขณะที่ฮิสโตแกรมสี่ไม่สามารถทำได้ และ จากการทดลองที่ 1 ถึงการทดลองที่ 4 พบว่า โครงข่ายนิเวศแบบ 1 ชั้นซ้อนที่ใช้ร่วมกับการวิเคราะห์ส่วนประกอบमुखสำคัญสามารถปรับปรุงความสามารถในการจำแนกประเภทกีฬา

จากการทดลองที่ 5 เป็นการวิเคราะห์ประสิทธิภาพของโครงข่ายนิเวศร่วมกับ การวิเคราะห์ส่วนประกอบสำคัญเมื่อจำนวนประเภทกีฬาจำนวนมากขึ้นพบว่า ออโตคอร์รีโลแกรมสามารถยังใช้ได้กับทุกประเภทกีฬา และ ความแม่นยำของการจำแนกที่ได้ยังอยู่ในช่วงที่ความ แม่นยำสูง จึงทำให้แน่ใจได้ว่า ออโตคอร์รีโลแกรมเป็นตัวแทนประเภทวิถีทัศนกีฬาได้เป็นอย่างดี

สำหรับการทดลองที่ 6 เป็นการทดลองกับตัวจำแนกในปริภูมิคู่ศูนย์พบว่า ออโต คอร์รีโลแกรมยังสามารถจำแนกประเภทกีฬาได้ความแม่นยำที่มากที่สุด นอกจากนี้ตัวจำแนกใน ปริภูมิคู่ศูนย์จำนวนพารามิเตอร์ที่ใช้ในหน่วยความจำน้อยกว่าโครงข่าย และ ขนาดของปริภูมิมี ขนาดเล็ก ในขณะที่มีประสิทธิภาพเทียบเท่ากับโครงข่ายนิเวศ และ พบว่าสำหรับฮิสโตแกรมสี และ ออโตคอร์รีโลแกรม เมื่อจำนวนประเภทกีฬาเพิ่มมากขึ้นจะให้ขนาดของปริภูมิคู่ศูนย์ลดลง เนื่องจากทำให้ เมทริกซ์การแปลงทำให้การกระจายของข้อมูลภายในกลุ่มอยู่ใกล้ยิ่งขึ้นทำให้ ขนาดของปริภูมิคู่ศูนย์ใช้จำนวนน้อยลง และ การเคลื่อนที่ของกล้อง และ ขอบสีพบว่าไม่มีปริภูมิคู่ ศูนย์ดังนั้นเมื่อเพิ่มประเภทกีฬาจึงขนาดปริภูมิคงที่เท่ากับขนาดของปริภูมิหลักเท่านั้น



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอการจำแนกวีดิทัศน์ประเภทกีฬาโดยใช้ลักษณะเฉพาะระดับล่างของวีดิทัศน์ ได้แก่ ฮิสโตแกรมสี ออโตคอรรีโลแกรม การเคลื่อนที่ของกล้อง และ ขอบสี นอกจากนี้ยังได้นำเสนอเทคนิคการจำแนกประเภทวีดิทัศน์ประเภทกีฬาด้วยการใช้การวิเคราะห์ส่วนประกอบमुखสำคัญร่วมกับโครงข่ายนิเวรอล และการจำแนกประเภทวีดิทัศน์ในปริภูมิผู้ศูนย์ ในงานวิจัยที่ใหญ่นิยมใช้ลักษณะเฉพาะของสี ได้แก่ ฮิสโตแกรมสี โดยที่จะให้ผลความแม่นยำในการจำแนกที่สูงที่เมื่อเปรียบเทียบกับลักษณะเฉพาะประเภทอื่นๆ เนื่องจากข้อมูลสีสามารถแสดงประเภทกีฬาที่แตกต่างกันได้ เช่น กีฬาฟุตบอลมักจะพบพื้นสนามฟุตบอลเป็นสีเขียว กีฬามวยไทยมักจะพบสีผิวของมนุษย์ เป็นต้น แต่ฮิสโตแกรมสีจะอธิบายเพียงการกระจายตัวของสีภายในภาพเท่านั้น ซึ่งจะจำแนกได้เฉพาะฉากที่มีพื้นสนามการแข่งขันเท่านั้นซึ่งให้ลักษณะการกระจายตัวของสีที่โดดเด่นสีใดสีหนึ่ง จึงไม่สามารถจำแนกฉากนั้นไปยั้งนักกีฬาหรือฉากของผู้ชมได้ ดังนั้นในวิทยานิพนธ์ฉบับนี้จึงได้ศึกษาและค้นหาลักษณะเฉพาะระดับล่างของวีดิทัศน์กีฬาที่สามารถนำมาจำแนกประเภทกีฬาที่มีการปะปนของฉากการแข่งขัน ฉากนั้นนักกีฬา และ ฉากผู้ชม นั่นก็คือลักษณะเฉพาะระดับล่างที่สามารถเป็นตัวแทนของประเภทกีฬาชนิดต่างๆ ได้

นอกจากนี้ยังได้ปรับปรุงเทคนิคการจำแนกประเภทวีดิทัศน์กีฬาด้วยโครงข่ายนิเวรอล โดยการนำการวิเคราะห์ส่วนประกอบमुखสำคัญซึ่งเป็นกรรมวิธีประมวลผลสัญญาณเบื้องต้นร่วมกับโครงข่ายนิเวรอล เนื่องจากโครงข่ายนิเวรอลที่ในการฝึกฝนเป็นเวลานานและใช้ขนาดหน่วยความจำในเก็บพารามิเตอร์ค่าถ่วงน้ำหนักเป็นจำนวนมากเพื่อให้ได้ผลความแม่นยำที่ดี จึงได้นำเสนออีกหนึ่งเทคนิคคือการใช้ตัวจำแนกประเภทวีดิทัศน์ในปริภูมิผู้ศูนย์ ซึ่งใช้พารามิเตอร์ในการคำนวณน้อยกว่าและฝึกฝนได้รวดเร็วกว่าโครงข่ายนิเวรอล

จากการศึกษาลักษณะเฉพาะระดับล่างของวีดิทัศน์แบบต่างๆ พบว่าออโตคอรรีโลแกรมเป็นลักษณะเฉพาะที่เหมาะสมกับการจำแนกประเภทกีฬามากที่สุด เนื่องจากสามารถให้ผลความแม่นยำของการจำแนกประเภทกีฬามากที่สุด ทั้งนี้เพราะว่าออโตคอรรีโลแกรมเป็นลักษณะเฉพาะของสีที่สามารถอธิบายสหสัมพันธ์เชิงพื้นที่ของสีเดียวกันได้ จึงสามารถแยกแยะที่

เป็นฉากเน้นตัวนักกีฬา และ ฉากผู้ชม ที่อยู่ในกีฬาต่างประเภทได้ ซึ่งจากการทดลองพบว่าเซตของระยะห่างระหว่าง 2 จุดภาพที่เหมาะสมกับการจำแนกประเภทกีฬามากที่สุดเป็น {1,3,5,7,9} เมื่อโครงข่ายนิรอรอลเป็นแบบ 1 ชั้นซ่อนที่มีจำนวนหน่วยชั้นซ่อนเป็น 24 หน่วยนิรอรอลและขนาดของปริภูมิส่วนประกอบมุขสำคัญเป็น 240 มิติ สำหรับ 7 ประเภทกีฬาได้แก่ บาสเก็ตบอล มวยไทย ฟุตบอล กอล์ฟ เทนนิส สปีงบอร์ด และ วอลเลย์บอล ซึ่งการฝึกฝนให้กับโครงข่ายนิรอรอลจะต้องหยุดฝึกฝนเมื่อค่าเฉลี่ยของความผิดพลาดจากข้อมูลวาลิเดต (Validate data) มีแนวโน้มสูงขึ้นได้ผลการทดลองที่ดีที่สุด จึงทำให้แน่ใจได้ว่าจะสามารถนำมาใช้งานทางระบบฐานข้อมูลวีดิทัศน์ประเภทกีฬา เช่น ฐานข้อมูลภายในสถานีโทรทัศน์ เพื่อทำดัชนีที่ใช้ความหมายตามประเภทของวีดิทัศน์ หรือ การให้คำประกอบวีดิทัศน์กีฬา เป็นต้น

5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

งานวิจัยนี้สามารถนำไปศึกษาและประยุกต์การใช้งานในระบบการให้คำประกอบวีดิทัศน์กีฬา หรือ ตรวจสอบเหตุการณ์กีฬาได้ โดยลักษณะเฉพาะระดับกลางที่น่าเสนอนี้สามารถปรับปรุงให้มีประสิทธิภาพมากยิ่งขึ้นได้ ประการแรก การใช้การรู้จำแบบรูปเทคนิคอื่นๆ เช่น การตัวจำแนกหลายตัวจำแนก แล้วทำการประมวลผลภายหลัง (post-processing) เพื่อตัดสินประเภทกีฬาอีกครั้งหนึ่ง โดยการเลือกประเภทกีฬาจากตัวจำแนกหลายตัวให้คะแนนเสียง ผลการจำแนกได้จากการให้คะแนนเสียงมากที่สุดเป็นประเภทของกีฬา ประการที่ 2 เมื่อจำนวนประเภทวีดิทัศน์กีฬามากขึ้นอาจนำลักษณะเฉพาะระดับกลางที่ได้นำเสนอมาใช้ร่วมกันเพื่อให้ได้ประสิทธิภาพการจำแนกมากที่สุด เป็นต้น

ประการที่ 3 ออโตคอรรีโลแกรมเป็นปริภูมิที่มีขนาดใหญ่ อาจจะใช้การลดจำนวนมิติเพื่อปรับปรุงประสิทธิภาพของโครงข่ายนิรอรอลได้อีกเนื่องจากโครงสร้างนิรอรอลไม่จำเป็นต้องใช้โครงข่ายที่ใหญ่เพื่อต้องการ Generalization ที่ดีสำหรับการจำแนกประเภทวีดิทัศน์กีฬา ประการสุดท้าย ลักษณะเฉพาะอื่นๆ เช่น ลักษณะการเคลื่อนไหวที่กลิ้ง หรือ ขอบสีที่น่าเสนออาจจะนำมาใช้ในการตรวจหาฉากเหตุการณ์อื่นๆ เช่น ฉากการแข่งขัน หรือ ฉากผู้ชมได้ ที่เกิดขึ้นในวีดิทัศน์กีฬาได้ ซึ่งจะต้องมีการทำการวิจัยกันต่อไป

รายการอ้างอิง

1. C. Colombo, A. Del Bimbo, and P. Pala. Retrieval of Commercial by video Semantics. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. (June 1998) : 572-577.
2. J. Assfalg, M. Bertini, C. Colombo, A. Del Bimbo. Semantic Annotation of Sports Videos. IEEE Multimedia Magazine. 9, 2. (April 2002) : 52-60.
3. C. Colombo, A. Del Bimbo and P. Pala. Semantic in Visual Information Retrieval. IEEE Multimedia Magazine. 6, 3. (July 1999) : 38-53.
4. X. Gibert, H. Li and D. Doermann. Sport video Classification using HMMs. In Proc. IEEE Conf. On Multimedia and Expo. 2. (July 2003) : 345-348.
5. D.H. Wang, Q. Tain, S. Gao and W.K. Sung. News Sport Video Shot Classification With Sport Play Field and Motion Features. In Proc. IEEE Conf. On Image Processing. 4. (October 2004) : 2247-2250.
6. L.Q. Xu and T. Li. Video Classification Using Spatial-Temporal Feature and PCA. In Proc. IEEE Conf. On Multimedia and Expo. 3. (July 2003) : 485-488
7. Y.F. Ma and H.J. Zhang. Motion Pattern Based Video Classification Using Support Vector Machines. In Proc. IEEE Intenational Symoisium on Circuit and Systems. 2. (May 2002) : II-62 – II 72.
8. K. Messer, W. Christmas and J. Kittler. Automatic Sports Classification. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. 2. (August 2002) : 1005-1008.
9. D. Sadlier and N.E. O'Connor. Event Detection Based on Generic Characteristics of Field-Sports. In Proc. IEEE Conf. On Multimedia and Expo. (July 2005) : 759-762.
10. S. Takagi, S. Hattori K. Yokoyama, A. Kodate and H. Tominaga. Sport video categorizing method using camera motion parameters. In Proc. IEEE Conf. On Multimedia and Expo 2. (July 2003) : 461-464.

11. E. Jaser, J. Kittler and W. Christmas. Hierarchical Decision Making Scheme for Sports Video Categorisation with Temporal Post-Processing. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition 2. (July 2004) : 908-913.
12. J. Assfalg, M. Bertini, C. Colombo, A. D. Bimbo and W. Nuniati. Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation. In Proceedings of the 2003 ACM symposium on Applied computing. (2003) : 769-773.
13. J. Assfalg, M. Bertini, C. Colombo and A.D. Bimbo. Feature extraction and content analysis for sports videos annotation. In Proceedings of the 2001 ACM workshops on Multimedia information retrieval. (2001) : 60-65.
14. M. Bertini, A. D. Bimbo, R. Cucchiara and A. Prati. Semantic video adaptation based on automatic annotation of sports videos. In Proceedings of the 6 th ACM SIGMM international workshop on Multimedia information retrieval. (2004) : . 291-298.
15. M. Bertini, A. D. Bimbo and W. Nunziati. Player identification in soccer videos. In Proceeding of the 7 th ACM SIGMM international workshop on Multimedia information retrieval. (2005) : 25-32.
16. L.Y.Duan, M. Xu. T.S. Chua, Q. Tain and C.S. Xu. A mid-level representation framework for semantic sports video analysis. In Proceeding of 11 th ACM international conference on Multimedia. (2003) :33-44.
17. Y. Rui, Thomas S. Huang. Handbook of Image and Video processing. Academic press. : 706-707
18. M. S. Lew. Principles pf Visual Information Retrieval. The Netherland : Institute of Advanced Computer Science, Leiden University, 2001.
19. I. Koprinska. Temporal Video segmentation: A Survey. Institute for Information Technologies. Bulgaria
20. F. Idris, S. Panchanathan. Review of Image and Video Indexing. Journal Visual communication and image representation 8, 2. (1997) :146-166.
21. A. Hajalic. Shot-Boundary Detection: Unraveled and Resolved?. IEEE Trans. on Circuits and Systems for Video Technology. 12, 2. (February 2002) : 90-105.

22. A. Hanjalic and H. J. Zhang. An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis. IEEE Trans. on Circuit and Systems for Video Technology. 9, 8. (December 1999) : 1280-1289.
23. Y. Zhuang, Y. Rui, T. S. Huang and S. Mehrotra. Adaptive key frame extraction using unsupervised clustering. In Proc. IEEE Conf. On Image Processing. 1, (October 1998) : 866-870.
24. H. J. Zhang, J. Wu, Di. Zhong and S. W. Smolair. An integrated system for content-based video retrieval and browsing. Pattern Recognition. 30, 4, (1997).
25. W. Wolf. Key frame selection by motion analysis. In Proc. IEEE Conf. On Acoustics, Speech, and Signal Processing. 2. (May 1996) : 1228 – 1231.
26. J. Shi, M.A. Nascimento and T. Wang. Experimental Results Towards Content-Based Sub-Image Retrieval. Department of Computing Science University of Alberta, Canada.
27. ภาสวรรณ สกุลสุจิราภา. การพัฒนาเทคนิคการค้นคืนภาพดิจิทัลโดยการใช้อัตราวิดีโอแถมและการแปลงเวฟเล็ต. วิทยานิพนธ์ปริญญาโทบริหารบัณฑิต สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2547.
28. J.J. Barron, D.J. Fleet, S.S. Beauchemin and T.A. Burkitt. Performance of Optical Flow Techniques. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. (June 1992) : 236-242.
29. R. Fablet and P. Bounthemy. Statistical Motion-based object Indexing using optical flow. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. 4, (September. 2000) : pp. 287 – 290.
30. N. Vasconcelos and A. Lippman. A Spatiotemporal Motion model for Video Summarization. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. (June 1998) : 361 – 366.
31. N. Dimitova, and F. Golshani. Motion Recovery for Video Content Classification. ACM Trans. Information System. 3, 4 (1995) : 408-439.
32. L. Wei, X. Liu, S. Lin, G. Xu and H-Y. Shum. Generic Slow-Motion Replay Detection in Sports Videos. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. 3. (October 2004) : 1585-1588.

33. R.C. Gonzalez and R. E. Woods. Digital Image Processing. 2 nd ed. New Jersey, United States of America : Prentice hall, 2001.
34. J. Fan, D. K. Y. Yau, A.K. Elmagarmid, and W. G. Aref. Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing. IEEE Trans. On Image processing. 10. 10. (2001).
35. A. Amato, and V.D. Lecce. Edge Detection Techniques in Image Retrieval: The semantic meaning of edge. In Proc 4 th EUASIP Conference focused on Video/Image Processing and Multimedia Commuication. (2003).
36. W.H. Adam, G. Iyengar,C.-Y. Lin, M. R. Naphade, C. Neti, H. J. Nock, and J. R. Smith. Semantic indexing of multimedia content using visual audio, and text cues. in EURASIP JASP. 2, (2003) :. 170-185.
37. C. Colombo, A. Del Bimbo and P. Pala, "Retrieval of Commercial by video Semantics," In Proc. IEEE Conf. On Computer Vision and Pattern Recognition. (June 1998) : 572-577.
38. W. Zhou, S. Dao, C.-C. and Jay Kuo. On-line knowledge-and rule based video classification system for video indexing and dissemination. Information Systems. 27, 8 (Dec 2002) : 559-586.
39. W. Zhou, A. Vellaikal, and C. Kuo. Rule-based video classification system for basketball video indexing. ACM Multimedia. 18. (2000), : 128-132.
40. J. Kiltter, M. Messer, W. J. Christmas, B. Levienaise-Obadia, and D. Koubaroulis. Generation of Semantic Cues for Sport Video Annotation. In Proc. IEEE Conference on Image Processing. 3, (Oct. 2001) : pp. 26-29.
41. S. Kumar. Neural Networks: A Classroom approach. : McGraw-Hill, 2005.
42. R.O. Duda, P.E. Hart and D. G. Stork. Pattern Classification. : John Willey & Son, 2001.
43. X.Qui and L.Wu. Null Space-based with LDA weighted dual personal subspaces for face recognition. In Proc. IEEE Conference on Image Processing. 2. (September 2005) :934- 937.
44. W. Liu, Y. Wang, S. Z. Li, and T. Tan. Null Space Approach of Fisher Discriminant Analysis for Face Recognition. In Proc. of the 8 th European Conference on Computer Vision (ECCV), Biometric Authentication Workshop. (2004).

45. N. Vaswani and R. Chellappa. Classification probability analysis of principal component null space analysis. In Proc. IEEE Conference on Pattern Recognition. 1 (August 2004) : 240-243.
46. J. Yang, H. Yu, and W. Kunz. An Efficient LDA Algorithm for Face Recognition. In Proc. of the 6 th Inter. Conf. on Control, Automation Robotics and Vision, (2000).
47. K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, 2 nd ed, 1990.
48. P. L. Bartlett. The sample complexity of pattern classification with neural networks : the size of the weights is more important than the size of the network. IEEE Trans on Inforamtion Theory. 44. 2, (March 1998) : 525-536.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทความทางวิชาการที่ได้รับการเผยแพร่

1. N. Watcharapinchai, S. Aramvith, S. Siddhichai, and S. Marukatat. Sports Video Classification Based On PCA of Color Histogram using Neural Network. In the International Workshop on Advanced Image Technology 2006. (IWAIT 2006, Okinawa, japan), 9-10 January 2006, pp 404-407.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

S12-2 SPORTS VIDEO CLASSIFICATION BASED ON PCA OF COLOR HISTOGRAM USING NEURAL NETWORK

N. Watcharapinchai[†], S. Aramvith[†], S. Siddhichai^{}, and S. Marukatai^{*}*

[†]Digital Signal Processing Research Laboratory
Department of Electrical Engineering, Faculty
of Engineering
Chulalongkorn University, Bangkok 10330,
Thailand
Tel: (66-2) 218-6909, Fax: (66-2) 218-6912
Email : supavadee.a@chula.ac.th

^{*}National Electronics and Computer
Technology Center (NECTEC)
Thailand Science Park, Pathumthani 12120,
Thailand
Tel: (66-2) 564-6900, FAX: (66-2) 564-6873
Email : yai@nectec.or.th

ABSTRACT

We investigated the problem of automating sports video classification by analyzing the low-level visual signal patterns. In this paper, a sports video classification system using multilayer perceptron neural network is proposed. The classification feature is based on principal component analysis of color histogram in RGB Spaces in order to represent the field color ratio of video frames. The classification module is based on a multilayer perceptron. Four different kinds of popularly televised sports are studied, namely Thai boxing, football, tennis, and volleyball. The experiments were emphasized on classifying video sequence at frame level. Experimental results indicated that our classification system is suited for sport video classification.

1. INTRODUCTION

Presently the advances in computing, network, and multimedia technology mandates the development of automatic video content classification and categorization possible for efficient content-based video searching and retrieval system. This is particularly important for the management and access of large digital video library and broadcasting video materials of TV/cable broadcasting stations. The basic approach of automatic video content classification is to extract low level video features such as color, shape, edge, or motion, etc., and to relate those features to their semantic meanings. Yet this is still the challenging issues for researchers since there exists the large gap between those two entities.

The rest of this paper is organized as followed. Section 2 review some related works. The details of color histogram feature are briefly described in Section 3.1. An introduction to Principal component analysis is provided in Section 3.2. Neural network theory is briefly described in Section 3.3. The experimental results are discussed in Section 4. Finally, conclusions of this work are drawn in Section 5.

2. RELATED WORKS

There has been a vast interest related to sports video classification. Generally, the sports video will be classified into several predefined classes and indexed semantic definitions or by some particular sport events. Some researchers constructed the model to classify the events in football [8] or basketball [9], while several others proposed

general classifier for several kinds of sports [1-4,7]. For the latter cases, we briefly summarized the work related to this area.

Gilbert et al. [1] proposed techniques in sports video classification using Hidden Markov Models (HMMs) to model each sports genre with motion and color features.

Xu and Li [3] used Gaussian Mixture Models (GMMs) to classify 5 TV broadcasting: sports, cartoon news, commercial, and music. In this work principal component analysis (PCA) is used to extract meaningful feature from texture color and motion features.

The above 2 works are based on generative model of each sport or video genre. Adding new genre to the system require building the new model for this genre without touching existing model. However, these generative models are not optimized for classification task and other discriminative approaches could be used, especially if the number of genres is fixed and unchanged.

Wang et al. [2] proposed sports news shot classification techniques C4.5 decision tree. The feature vectors were 11 dimensions which included field color ratio in frame, motion feature vector, and ratio of movement background.

Assafalg [4] proposed the use of neural network as a classifier for sports video by using HSV color histogram feature and edge intensity histogram feature.

Ma and Zhang [7] proposed a semantic classification scheme based on motion texture pattern using Support Vector Machines (SVMs). SVMs is based on a firm statistical theory and obtains very good result on many classification tasks. However, we are unable to control the number of support vectors, thus the resulting model. Therefore, the resulting model may be very large, hence slow down the running time.

3. THE PROPOSED CLASSIFICATION SYSTEM

In this paper, we proposed a classification system based on the neural network. The visual feature used is color ratios from color histogram. We apply the PCA to extract meaningful features as well as reduce the dimension of histogram features. The block diagram of our proposed scheme is shown in Fig. 1.

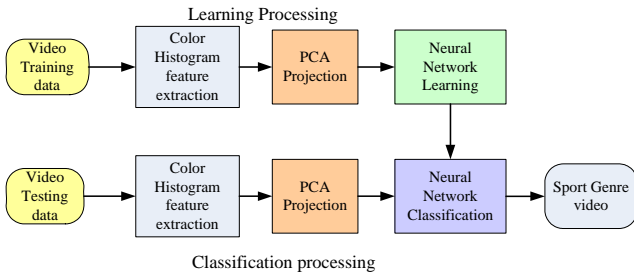


Figure 1 Sports video classification diagram.

In the following subsections, we describe the color histogram features in 3.1, the PCA feature extractor in 3.2 and the Neural network in 3.3.

3.1 Color histogram feature

The color histogram is one of the most important and simplest techniques for image and video indexing and retrieval system. The computation is simple yet it is efficient and effective. In order to compute the color histogram of a frame in video sequence, the color model of each frame must be quantized into n colors to reduce complexity of the computation, as shown in Eq. (1).

$$H(M) = \langle h_1, h_2, \dots, h_n \rangle \quad (1)$$

The vector $H(M)$ is an n -dimensional feature vector representing the color histogram for a frame M of the video sequence and h_j is the number of pixels as color j divided by the total number of pixel in a frame for arbitrary sized frames. The computational complexity of color histogram method is directly related to the dimension of the feature vector $O(N^2)$ where N is the number of pixels.

3.2 Principal Component Analysis

The analysis of a high-dimensionality data is often a complex task. Classification systems are adversely affected by high dimensionality. To handle this problem, classical approach consists in reducing or extracting the small set of meaningful features from the original features.

PCA allows achieving this goal by finding the linear transformation of the original features. The new, transformed features, are orthogonal and can be sorted by their importance. This allows extracting most meaningful features from the original features. Therefore reducing the dimension of feature vector before passing to the classifier. The linear transformations corresponding to the most meaningful transformed features are called the principal components or principal axes. We will use the projection of histogram features onto the principal subspace as the classification features.

PCA is performed on a training database, which includes the video feature of all classes to be identified. Given $\mathbf{x}_i, i = 1, 2, \dots, N$ as the feature vector is n -dimension feature including N sample data, the mean vector and covariance matrix are given, respectively, by

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (2)$$

$$\mathbf{C} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (3)$$

Then the P principal components correspond to the first P significant eigenvector of \mathbf{C} . Let $\mathbf{v}_i, i = 1, 2, \dots, P$ be these principal components and define the following matrix $\mathbf{U} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_P]$ of $n \times P$ dimension. An arbitrary n -dimension original feature vector \mathbf{x} can be represented as new compact feature space as shown in Eq. (4)

$$\mathbf{y} = \mathbf{U}^T (\mathbf{x} - \mathbf{m}) \quad (4)$$

3.3 Neural network

The artificial Neural Network is a computation model which tries to mimic human brain's functionalities. It is now a classic classification tools that has been used widely, such as optical recognition, speech recognition, regression, and pattern recognition [4]. Because of neural network can endure from noise, distorted pattern, and incompletely practical sample data.

In the forward pass, the weights remain unaltered throughout the network, and the function signals of the network are computed on a neural-by-neural basis. Specification, the function signal appearing at output of neural j is computed as Eq. (5)

$$o_j(n) = f(\text{net}_j(n)) \quad (5)$$

Where $\text{net}_j(n)$ is the net internal activity level of neural j , define by Eq. (6)

$$\text{net}_j(n) = \sum_{k=0}^p w_{kj}(n) o_k(n) \quad (6)$$

Where p is the total number of inputs, excluding the threshold, applied to neural j , and $w_{kj}(n)$ is the weight connecting neural k to neural j , and $o_k(n)$ is the input signal of neural j .

If the neural j is in the first hidden layer, then the index k refers to the k th-input terminal of the network. On the other hand, the neural j is in the output layer, the j th-output terminal of the network.

Where $f(\text{net}_j)$ is the activation transfer function in the neural j . The continuous nonlinear activation function commonly used in multilayer is the sigmoid transfer function or hyperbolic tangent transfer function.

In backward pass, the links between layers are weight parameters that can be adjusted by the Back-propagation Learning Algorithm as shown in Eq. (7)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta \mathbf{w}(n) \quad (7)$$

Where $\mathbf{w}(n+1)$ is a next weight vector, $\mathbf{w}(n)$ is a previous weight vector, and $\Delta \mathbf{w}(n)$ is different weight vector, determined by Eq. (8) and (9).

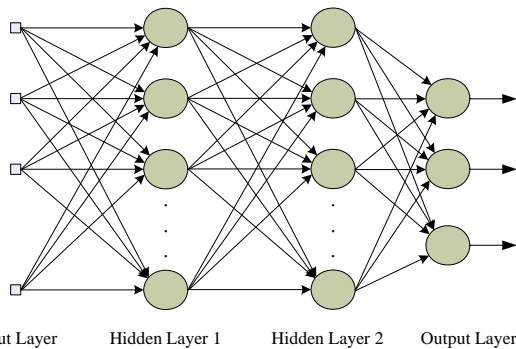


Figure 2 shows a simple feed-forward multilayer neural network that set a classifier. The connection of network distinguishes 3 layer connections: Input Layer, Hidden Layer and Output Layer.

$$\Delta w_{ij} = \eta \delta_j o_i \quad (8)$$

$$\delta_j = \begin{cases} f'(net_j)(t_j - o_j) & \text{if } j \text{ is output unit} \\ f'(net_j) \sum_k \delta_k w_{jk} & \text{if } j \text{ is hidden unit} \end{cases} \quad (9)$$

The output of the network o_j is compared with the desired response t_j , obtaining the error signal for adjusting the new weight vector.

In this work, we use 3 layer feedforward network with the number of input nodes equal to the dimension of principal subspace. The number of output nodes is set to the number of sport genres. The number of hidden nodes is adjusted experimentally.

Each output of neural network estimates a posterior probability of sports genre. A sports genre was chosen based on Bayes decision rule, choosing genre by maximum estimated output unit as shown in Eq. (10)

$$o_k(\mathbf{x}) = \max_j (o_j(\mathbf{x})) \quad (10)$$

4 EXPERIMENTAL RESULT

The video sequences used in the experiment were obtained from Thai television broadcasting recorded in MPEG-1 format with resolutions of 448x384 pixels.

4.1 Experimental Data

In the experiment, the test data were obtained from TV Broadcast of 14 video sequences with 4 sport genres namely Thai boxing, football, tennis, and volleyball, in total 12 hours of video for training and test data. A 64-bin RGB histogram is extracted from 100 video frame (4 seconds). Table 1 shows the number of sequence of each sport genres, the number of training and test frames.

Table 1 The number of Train-Test sample Data

	#Videos	#Training frames	#Test frames
Thai Boxing	3	600	300
Football	4	600	300
Tennis	5	600	300
Volleyball	2	300	100
Total	14	2000	1000

4.2 The number of PCA dimension and of histogram bin

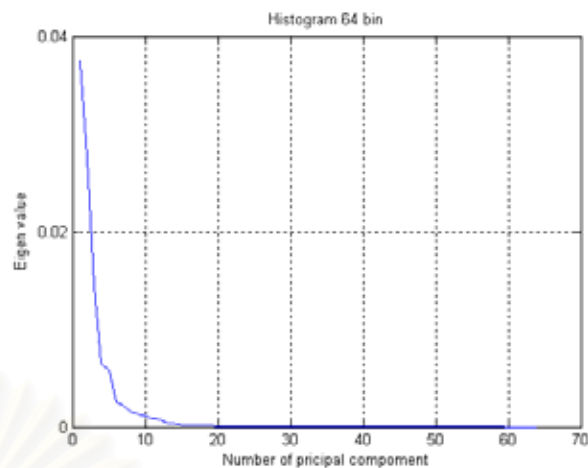


Figure 3 : Eigenvalue on PCA dimension in 64 bin

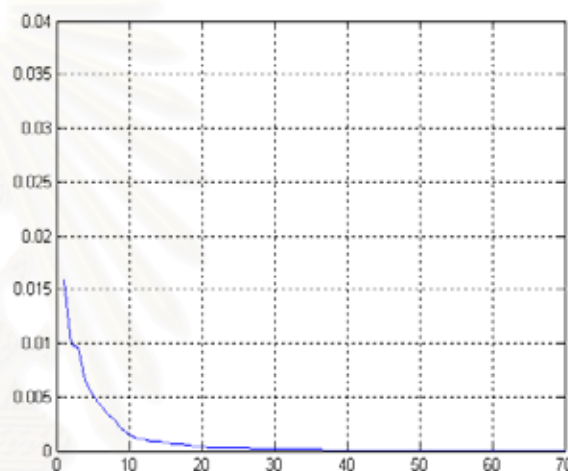


Figure 4 : Eigenvalue on PCA dimension in 512 bin

By projecting all sample data, both training and test data, of color Histogram feature into a pre-trained PCA space, we obtain a compact PCA feature vector. In Fig 3 and 4, it can be observed that eigenvalue or variant in each principal component axis of the 64-bin color histogram and 512-bin color histogram, respectively.

From Figs. 3 and 4, both of the number of PCA dimension increase then the eigenvalue decrease and the dimension more than 20 dimension will be converged to zero values. In the zero eigenvalue in the PCA dimension axis can not be used to represent the data.

In order to learn from training data, we randomly subdivided the training data into two sets: learn, and validate set. The ratio between learning and validation data sets in training data is 2 to 1. The learning processing will be stopped when there is an increasing errors of validate data set. In the other words, this is done to improve generalization ability of neural network as too many iterative learning may results in the overfitting. The error of validate data set is mean square error (MSE) between output of network in forward pass and desired output.

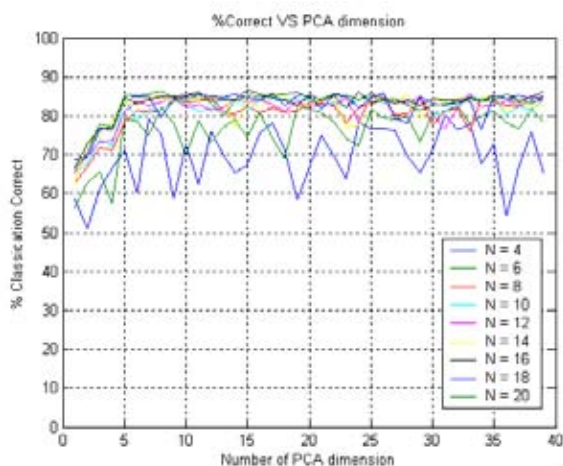


Figure 5 : Sport frame classification accuracy on the number of PCA dimension

We construct a 3-layer neural network, one input layer, one hidden layer, and one output layer. The activation transfer function in the hidden units was specified as Hyperbolic tangent transfer function and the output units set as Sigmoid transfer function. Given a limited training set, using too many hidden nodes can result in overfitting thus dropping the classification accuracy. In this work, the number of hidden units was varied between 4 and 40 units. The accuracy of experimental results according to different parameters is shown in Figure 5.

The results in Figure 5 shows that as the number of PCA dimension increased past 10 dimensions, it can be observed that the average classification performance was 85%. The best classification accuracy from 10-dimension PCA was achieved with 16 hidden units as shown in Table 2.

Table 2 Confusion Matrix of 10 PCA dimension

	Boxing	Football	Tennis	Volley
Boxing	95.66	0.33	3.33	0.66
Football	2.66	86.00	3.33	8.00
Tennis	5.00	4.33	87.66	3.00
Volley	10.00	3.00	7.00	80.00

5. CONCLUSION

In this paper, we have presented a sports video classification method using multilayer perceptron based on principal component analysis of color histogram. Our classification scheme allows classification of different sports genres. Experimental results indicate that PCA color histogram from video frames can provide very useful information for classification. The PCA can efficiently extract useful features from color histogram in sports video and reduce computational complexity in neural network. Also it has indicated that multilayer perceptron can effectively learn to bridge the low level features and semantic contents. We believe this can contribute to semantic content understanding in video.

6. ACKNOWLEDGEMENT

This research is supported by Thailand Graduate Institute of Science and Technology (TGIST), National Science and Technology Development Agency (NSTDA) and Ratchadapisek Somphot Endowment, Chulalongkorn University.

7. REFERENCES

- [1] X. Gibert, H. Li and D. Doermann, "Sport video Classification using HMMs," *IEEE ICME'03*, 2003.
- [2] D.H. Wang, Q. Tain, S. Gao and W.K. Sung, "News Sport Video Shot Classification With Sport Play Field and Motion Features," *IEEE ICIP'04*, 2004.
- [3] L.Q. Xu and T. Li, "Video Classification Using Spatial-Temporal Feature and PCA," *IEEE ICME'03*, 2003.
- [4] S. Kumar, "Neural Network: A Classroom approach," *McGraw-Hill*, 2005.
- [5] R.O. Duda, P.E. Hart and D. G. Stork, "Pattern Classification," *John Willey & Son, Inc*, 2001.
- [6] J.Assfalg, M. Bertini, C. Colombo, and A.D. Bimbo, "Semantic Annotation of Sport Videos," *IEEE Multimedia'02*, 2002.
- [7] Y.F. Ma and H.J. Zhang "Motion Pattern Based Video Classification Using Support Vector Machines," *IEEE ISCAS'02*, 2002.
- [8] J. Assfalg, M. Bertini, C. Colombo, A. D. Bimbo and W.Nunziati, "Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation," *ACM'03*, 2003.
- [9] W. Zhou, A. Vellaikal and C.-C Jay. Kuo, "Rule-based Video Classification System for Basketball Video Indexing," *ACM'02*, 2000.

ประวัติผู้เขียนวิทยานิพนธ์

นายณัฐชัย วัชรภินชัย เกิดวันที่ 30 พฤศจิกายน พ.ศ. 2524 ที่ จังหวัด สงขลา ได้เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2542 ซึ่งสำเร็จการศึกษาวิศวกรรมศาสตรบัณฑิต (เกียรตินิยมอันดับ 2) สาขาวิศวกรรมอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2545 และ เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546 โดยได้รับทุนสนับสนุนในการทำงานวิจัยจากสถาบันบัณฑิตวิทยาศาสตร์และเทคโนโลยีไทย สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย