



รายงานการวิจัย

เรื่อง

การรู้จำท่าทางมนุษย์จากภาพส่วนบุคคลผ่านการวิเคราะห์การใช้พลังงานร่างกาย

Human Action Recognition from Personal Photos via Analysis of Body Energy
Expenditure

โดย

นศัพ์ชาณัณ ชินปัญชัณนะ

มหาวิทยาลัยธุรกิจบัณฑิตยั

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตยั

พ.ศ. 2555

ชื่อเรื่อง : การรู้จำท่าทางมนุษย์จากภาพส่วนบุคคลผ่านการวิเคราะห์การใช้พลังงานร่างกาย

ผู้วิจัย : นศพัชราณัฐ ชินปัญชรณะ สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิตย์

ปีที่พิมพ์ : พุทธศักราช 2555 สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์

แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิตย์

จำนวนหน้าวิจัย : 68 หน้า ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์

คำสำคัญ : การประมวลผลภาพ, พลังงาน, ท่าทางมนุษย์, การรู้จำท่าทาง, การจำแนกภาพ

บทคัดย่อ

งานวิจัยทางด้านกระบวนการรู้จำและแปลความหมายท่าทางมนุษย์ได้รับความนิยมเพิ่มขึ้นเป็นอย่างมาก เนื่องจากสามารถนำผลลัพธ์ที่ได้ไปประยุกต์ใช้งานทางด้านอื่นอย่างหลากหลาย ดังนั้นในงานวิจัยนี้จึงนำเสนอโมเดลใหม่ที่ใช้ในการจำแนกการใช้พลังงานจากท่าทางการเคลื่อนไหวของมนุษย์แปลงเป็นความหมายของภาพ โดยโมเดลการใช้พลังงานนี้ได้ใช้พื้นฐานของการคำนวณมาจากทฤษฎีจากท่าทางทางกลศาสตร์ของการเคลื่อนที่ รวมทั้งทางสรีระวิทยาของร่างกายมนุษย์เพื่อที่จะแทนความหมายจากท่าทางโดยใช้การวัดประสิทธิภาพในการจำแนกภาพดังนี้ โครงข่ายประสาทเทียมแบบหลายชั้น ทฤษฎีแผนผังการจัดระเบียบตัวเอง ทฤษฎีเครือข่ายแบบเบย์ และซัพพอร์ตเวกเตอร์แมชชีน จากการทดลองสามารถสรุปได้ว่าโมเดลการใช้พลังงานสามารถช่วยในการแปลความหมายของภาพได้อย่างดี

เลขทะเบียน	0220889
วันลงทะเบียน	5 ต.ค. 2556
เลขเรียกชื่อ	006.42
	๖4577
	[๑๖๕]

Title : Human Action Recognition from Personal Photos via Analysis of Body Energy Expenditure

Researcher : Nuchanun Chinpanthana **Institution** : Dhurakijpundit University.

Year of Publication : 2012 **Publisher** : Dhurakijpundit University.

Sources : Dhurakijpundit University Research Center.

Number of Pages : 68 Pages **Copyright** : Dhurakijpundit University.

Keyword : Image processing, Energy Expenditure, Human Action, Human Recognition
Image Classification

Abstract

Human action recognition for interpreting at a semantic level has been and still a highly interesting and important research topic. The research results are capable of analyzing human action that we visually perceive in many aspects. Therefore the research requires an effective and competent approach to accurately interpret human action. In this paper, we present a novel model called the Energy expenditure model for producing more semantic classification. The energy expenditure model is based on the fundamental concepts of biomechanics that human movement in different classes is likely to spend different amounts of energy. The energy expenditure model is classified by using Bayesian Network, Multiple Feedforward Neural Network, Self-organizing maps, and Support vector machines. Experimental results show that the proposed provides much more authentic meaning of human actions

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จลุล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และแรงผลักดัน รวมถึงติดตามงานของผู้วิจัยอย่างต่อเนื่อง และอีกหลายฝ่ายที่ได้ให้ความช่วยเหลือเอื้อเฟื้อข้อมูลที่เป็นประโยชน์ในการทดลองโปรแกรม บันทึกผลการทดลอง และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ผู้ทรงคุณวุฒิ ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิต ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วย

นศพัชชาณ ชินปัญชธนะ

เมษายน 2556

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	(1)
สารบัญตาราง	(3)
สารบัญรูปภาพ	(4)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์	4
1.3 สมมุติฐาน	4
1.4 ขอบเขตของการวิจัย	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	5
1.6 นิยามคำศัพท์	6
บทที่ 2 ทฤษฎีและงานวิจัย	7
2.1 ทฤษฎีเบื้องต้น	7
2.2 การประมวลผลภาพดิจิทัล	13
2.3 การเปรียบเทียบความคล้ายของภาพ	19
2.4 การใช้โมเดลในการสร้างการตัดสินใจ	32
2.5 การวัดประสิทธิภาพ	34
บทที่ 3 ขั้นตอนวิธีการดำเนินงานวิจัย	36
3.1 ขั้นตอนการเตรียมข้อมูล	36
3.2 ขั้นตอนการประมวลผล	46
3.3 ขั้นตอนการวัดประสิทธิภาพ	52
บทที่ 4 ผลการทดลอง	53
4.1 การกำหนดข้อมูลภาพ	53
4.2 ผลการทดลอง	54

บทที่ 5 สรุปผลการทดลอง	59
5.1 สรุปและอภิปรายผลการวิจัย	59
5.2 ข้อเสนอแนะ	60
บรรณานุกรม	62
ประวัติผู้วิจัย	69

สรุปผล

สารบัญตาราง

ตารางที่หน้า

2.1	การวัดประสิทธิภาพ	34
3.1	ตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลักในการให้ความหมายภาพ	38
3.2	แหล่งฐานข้อมูลของคำหลัก	41
4.1	ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี naïve-Bayes	54
4.2	ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Multilayer	55
4.3	ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี SOM	55
4.4	ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Similarity Matching	55
4.5	ตารางเปรียบเทียบผลลัพธ์การจำแนกข้อมูล	56

สารบัญรูปภาพ

รูปภาพที่	หน้า	
2.1	การประมวลผลแบบค้นคืนด้วยคุณลักษณะพีเจอร์ระดับต่ำ	8
2.2	ตัวอย่างขององค์ประกอบของภาพ	9
2.3	ตารางแสดงการวัดประสิทธิภาพ	10
2.4	การแก้กองค์ประกอบของภาพ	11
2.5	การทำตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน	13
2.6	การแสดงตำแหน่งของฟิกเซลบนภาพดิจิทัลด้วยเมทริกซ์	15
2.7	แบบจำลองของภาพด้วยระนาบ	15
2.8	การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์	17
2.9	เซลระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม	20
2.10	ตัวแบบของนิวรอนในคอมพิวเตอร์	21
2.11	สถาปัตยกรรมแบบป้อนไปข้างหน้า	22
2.12	สถาปัตยกรรมแบบมีการป้อนไปเวียนกลับ	22
2.13	โครงสร้างแบบป้อนไปข้างหน้าหลายชั้น	23
2.14	การเรียนรู้แบบมีการสอน	24
2.15	การเรียนรู้แบบไม่มีการสอน	25
2.16	โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง	28
2.17	การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน	29
3.1	ขั้นตอนการจำแนกกลุ่มความหมายภาพ	36
3.2	ภาพตัวอย่างที่ถูกแก้ด้วยคำหลัก	37
3.3	ส่วนย่อยของโครงสร้างลำดับชั้นของคำหลัก“dog” บน WordNet	39

3.4	ความสัมพันธ์ของโครงสร้างคำหลักบน WordNet	49
3.5	โปรแกรม LabelMe บนเบราว์เซอร์	42
3.6	การเลือกสัดส่วนของวัตถุบนภาพ	43
3.7	ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe	43
3.8	ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe	44
3.9	ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ	45
3.10	ตัวอย่างของคำหลักในการให้ความหมายวัตถุบนภาพ	45
3.11	ตัวอย่างความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ	47
3.12	ความสัมพันธ์ของวัตถุแทนด้วยแนวคิดกราฟ	48
4.1	แสดงความผิดพลาดของการเคลื่อนที่ของถุงปูนซีเมนต์บนสายพาน	52
4.2	แสดงการเปรียบเทียบผลการทดลองการตรวจนับการเคลื่อนถุงปูนซีเมนต์บนสายพานด้วยโปรแกรมประมวลผลภาพแบบแทมเพลตแมชชีง	54
4.3	ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ	57

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

สมัยก่อนการประมวลผลภาพ จะอยู่ในรูปของการประมวลผลสัญญาณแอนะล็อก (analog) โดยใช้อุปกรณ์จับภาพแต่งภาพที่ถ่ายออกมายังคงเป็นภาพขาวดำและไม่ชัดเจนเท่าที่ควร ทำให้ส่วนใหญ่ภาพถ่ายมักจะถูกถ่ายเฉพาะภาพเหตุการณ์ที่สำคัญหรือบุคคลสำคัญเท่านั้น เนื่องจาก อุปกรณ์ที่นำมาใช้ในการถ่ายภาพ หรือบันทึกภาพนั้นซับซ้อนและมีความยุ่งยากในการใช้งาน รวมทั้งราคาสูง แต่อย่างไรก็ตามได้มีการพัฒนาปรับปรุงเทคโนโลยีเพื่อตอบรับกับความต้องการที่เพิ่มขึ้น ทำให้อุปกรณ์ถ่ายภาพในปัจจุบันมีราคาถูกลง และสะดวกต่อการใช้งานมาก และพัฒนาต่อเนื่องไปเป็นอุปกรณ์ถ่ายภาพแบบดิจิทัล (digital) ทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง ด้วยเหตุนี้ทำให้เกิดปัญหาในการจัดเก็บข้อมูลภาพรวมถึงการค้นหาข้อมูลภาพ (image retrieval) ที่เพิ่มมากขึ้นอย่างต่อเนื่อง ทำให้การประมวลผลภาพดิจิทัล (digital image processing) เป็นสาขาที่ได้รับความนิยมในการวิจัยค่อนข้างมาก ทั้งทางด้านภาพ (still image) และ วีดีโอ (video) เพราะสามารถที่จะนำไปประยุกต์ได้หลายด้าน เช่น ทางด้านการแพทย์ ทางด้านอุตสาหกรรม หรือนำมาประยุกต์ใช้ในส่วนของการจัดการจราจร ในการตรวจจับทะเบียนรถยนต์ เป็นต้น

เมื่อเทคโนโลยีมีการพัฒนาไปอย่างรวดเร็ว โดยเฉพาะอย่างยิ่งอุปกรณ์ถ่ายภาพอิเล็กทรอนิกส์ ทำให้อุปกรณ์ถ่ายภาพ โดยเฉพาะภาพถ่ายส่วนบุคคล (personal images) มีเพิ่มมากขึ้น ทำให้การสืบค้นเพื่อให้ได้ภาพที่ต้องการทำได้ค่อนข้างลำบาก มีการใช้เทคนิคในการสืบค้นที่เริ่มจากการสกัดข้อมูลภายในภาพ (feature extraction) ออกมาเป็นส่วนๆที่เรียกว่า ข้อมูลระดับต่ำ (low-level features) [Qian Huang,1995][Vailaya A.,2001][C. Carson, 1999] เข้ามาใช้ในการค้นหาภาพ เพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ เป็นการนำคุณลักษณะของภาพต่างๆหรือฟีเจอร์ (feature) ที่ถูกสกัด (extraction) อย่างเช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) ด้วยวิธีการ อัลกอริทึมต่างๆ เพื่อใช้สำหรับการสืบค้นข้อมูล ความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดคุณลักษณะข้อมูลภายในภาพ (feature extraction) ออกมาเป็นข้อมูลที่สามารถคำนวณได้ ข้อมูลส่วนนี้มักจะถูกเรียกกันว่าข้อมูลภาพระดับต่ำหรือ low-level features โดยข้อมูลนี้จะถูกนำมาใช้สำหรับการสืบค้นข้อมูลภาพแต่อย่างไรก็ตามการวิจัยในส่วนแรกนี้ยังคงมีการปรับปรุงพัฒนาอย่างต่อเนื่อง แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจ

ความหมายของภาพแทน การสืบค้นแบบข้างต้นเรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะพิจารณาข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B,2002] และแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) ลงบนภาพ เช่น “building”, “car” , “ball” เป็นต้น และใช้ความหมายหรือคำหลักนั้นเพื่อทำการสืบค้นข้อมูลแทนเมื่อเปรียบเทียบผลลัพธ์ของการค้นคืนที่ค่อนข้างแม่นยำมากกว่าการใช้พีเจอร์ระดับต่ำ แต่อย่างไรก็ตามยังขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้ร่วมด้วยนั้นจะเป็นลักษณะใด เช่น การใช้เทคนิคหาความสัมพันธ์ของแท็กชื่อวัตถุบนภาพ [Benitez A.B,2002][R. Zhao,2002][Tele Tan, 2002] ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรม การใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ ที่ค่อนข้างขึ้นกับคำหลักที่ถูกให้ความหมายไว้ในภาพหรือ วัตถุบนภาพส่วนใหญ่ การวิจัยในส่วนแรกนี้ยังคงมีการปรับปรุงพัฒนาอย่างต่อเนื่อง เพื่อให้การสืบค้นข้อมูลภาพมีประสิทธิภาพและตรงตามความหมายของผู้ใช้งานมากที่สุด แต่อย่างไรก็ตามวัตถุประสงค์หลักของการทำความเข้าใจในภาพดิจิทัล โดยเฉพาะภาพถ่ายบุคคลด้วยแล้วนั้นจะต้องรู้ได้ว่า บุคคลในภาพนั้นกำลังจะทำกิจกรรมประเภทใดอยู่ จึงทำให้สามารถเข้าใจความหมายโดยรวมของภาพได้ ดังนั้นในงานวิจัยนี้จะมุ่งเน้นในการรู้จำกิจกรรมของมนุษย์ (recognition of human activities) เพื่อให้เป็นส่วนหนึ่งในการบอกความหมายของภาพโดยรวมได้

ปัจจุบันแอปพลิเคชันในการรู้จำกิจกรรมมนุษย์เป็นสิ่งที่จำเป็นและสามารถพบได้ทั่วไป [Ikizler N. 2009] เช่น วิดีโอตรวจจับ (video surveillance) วิดีโอคำอธิบายประกอบ (video annotation) การปฏิสัมพันธ์ระหว่างมนุษย์ (human interaction) และ การสืบค้นภาพ (image retrieval) [Ballan, L. 2009] จะเป็นการตรวจจับบุคคลที่กระทำพฤติกรรมที่ผิดปกติ เพื่อเป็นสัญญาณเตือนระวังอันตราย ทำให้ลดความผิดพลาดที่เกิดขึ้นจากการกระทำของมนุษย์ Ikizler N. [Ikizler N. 2009] ได้สร้างเทคนิคเพื่อรู้จำท่าทางเฉพาะ ท่าทางเดิน ท่าทางวิ่ง ด้วยวิธีการ Pose-based แต่ในความเป็นจริงแล้วการวิเคราะห์เพื่อรู้จำเพียงแค่สองท่าทางยังไม่เพียงพอ Lazebnik S. [Lazebnik S., 2006] ได้วิเคราะห์ท่าทางการกระทำของมนุษย์ จากการพิจารณาภาพพื้นหลังและวัตถุภายนอก ร่วมกันกับข้อมูลท่าทางเพื่อรู้จำท่าทางมนุษย์ เช่น ซี่มา คุยโทรศัพท์ หรือ ทำงาน โดยการวิเคราะห์ใช้หลักเกณฑ์ว่า การแสดงพฤติกรรมของมนุษย์ หรือการแสดงท่าทางเพื่อบอกกิจกรรมที่กระทำแบบเดียวกันมักจะมีการใช้วัตถุที่เกิดร่วมกันและสภาพของสิ่งแวดล้อมภายนอกคล้ายกันเสมอ เช่น การเล่นที่มีการใช้เครื่องมือร่วมที่พบบ่อยๆ การทำงานที่มีการใช้เครื่องคอมพิวเตอร์ โดยใช้การหาพื้นที่บริเวณโดยรอบจากภาพพื้นหลัง อาจจะเป็น ท้องฟ้า ถนน อาคาร หรือภาพที่เป็น ภายในบ้าน เป็นต้น แต่อย่างไรก็ตามการวิเคราะห์โดยการหาบริเวณด้วยวิธีการสกัดข้อมูลภาพระดับต่ำ (feature extraction with low-level) ยังเป็นวิธีการที่ต้องมีการทำวิจัยอยู่มาก นักวิจัยบางกลุ่มจึงพยายามใช้วิธีการแท็กข้อมูล (annotation) ลงบนภาพเพื่อให้การทำงานสมบูรณ์มากที่สุด แต่อย่างไรก็ตามการ

แท็กบนภาพยังไม่เพียงพอสำหรับความหมายของทั้งภาพ Marsza lek M. [Marsza lek 2009] ได้ใช้การย้อนกลับ (relevance feedback) ของกลุ่มภาพเพื่อให้ได้ความหมายของ กิจกรรมของมนุษย์มากที่สุด จากกลุ่มนักวิจัย [Li L. 2007][Gupta, A., 2009][Kjellström H. 2008][Lazebnik S. 2006][Bangpeng Y. 2010] ได้ทำการวิเคราะห์พฤติกรรมและการรู้จำท่าทางมนุษย์ที่เรียกว่า Spatial interaction model โดยใช้การวิเคราะห์จากความสัมพันธ์ระหว่างร่างกายกับวัตถุภายนอก การแทนความสัมพันธ์ที่เกิดขึ้นจะใช้ลักษณะของ High-level semantic interpretation ด้วยคำศัพท์ที่ถูกต้องแท็กบนภาพ เช่น on top, above, below, next to, near, far, overlap เป็นต้น จะเห็นว่าการสร้างโมเดลเน้นการปฏิสัมพันธ์ของวัตถุบนภาพ เป็นส่วนหลักและถูกวิเคราะห์ด้วยโมเดลเบย์ ซึ่งในความเป็นจริงแล้วการแสดง ความหมายของภาพ ส่วนใหญ่จะเกิดจากท่าทางของร่างกายมนุษย์มากกว่าสภาพแวดล้อม แต่อย่างไรก็ตามการสร้างระบบเพื่อการจำแนกและรู้จำกิจกรรมมนุษย์เป็นปัญหาที่ท้าทายเนื่องจาก สภาพแวดล้อม ภาพพื้นหลัง และ แสงเงา จะมีผลกระทบต่อกระบวนการเรียนรู้เพื่อทำการจำแนกและรู้จำภาพกิจกรรม ดังนั้นในงานวิจัยนี้จึงนำเสนอรูปแบบใหม่ของการรู้จำท่าทางมนุษย์ โดยนำหลักการของการใช้พลังงาน (energy expenditure) การเคลื่อนไหวของท่าทางร่างกาย (human action) เข้ามาวิเคราะห์ท่าทางมนุษย์ เพื่อสามารถนำไปใช้กับการแปลความหมายภาพ และการสืบค้นข้อมูลภาพได้อย่างมีประสิทธิภาพมากที่สุด

จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอการปรับปรุงวิธีการรู้จำท่าทางมนุษย์ (human action recognition) เพื่อให้ได้ผลลัพธ์ของพฤติกรรมมนุษย์ที่อยู่บนภาพอย่างแท้จริง โดยใช้วิธีการสกัดข้อมูลจากท่าทางการเคลื่อนไหวของร่างกายมนุษย์ (Human action movement) ด้วยโมเดลการใช้พลังงาน (Energy Expenditure) และทำเปรียบเทียบการรู้จำท่าทางทั้งหมด 3 วิธีการ[T. Mitchell 1997] โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network), แผนผังการจัดระบบตัวเอง (Self-organizing maps) และ เครือข่ายแบบเบย์ (Bayesian Network)

1.2 วัตถุประสงค์

1. เพื่อค้นหาเทคนิคใหม่ที่สามารถนำมาใช้ในการรู้จำภาพกิจกรรมมนุษย์
2. เพื่อสร้างโมเดลรูปแบบใหม่ในการแทนความหมายของร่างกายมนุษย์
3. เพื่อสร้างแบบจำลองวิเคราะห์การค้นหภาพรูปแบบใหม่
4. เพื่อพัฒนาเทคนิคในการหาความเหมือนกันของภาพ
5. สามารถนำไปประยุกต์กับการสืบค้นภาพได้

1.3 สมมติฐาน

งานวิจัยนี้จึงเสนอวิธีการแปลความหมายภาพด้วย โครงสร้างของท่าทางร่างกายมนุษย์ (human body action) โดยใช้หลักทฤษฎีการใช้พลังงาน (energy expenditure) จากการท่าทางมนุษย์สามารถบอกถึงความหมายของภาพมนุษย์ได้ โดยทำการเชื่อมโยงประสานกับวัตถุนภาพให้สอดคล้องตามความหมายของกระบวนการคิดแปลความหมายภาพของมนุษย์

1.4 ขอบเขตของการวิจัย

1. ข้อมูลภาพที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัล เฉพาะภาพธรรมชาติ (natural images) เช่น ภาพทิวทัศน์ ชายทะเล ภูเขา หรือ สวนสาธารณะ ภาพภายในอาคาร ภายในบ้าน สำนักงาน หรือ โรงงาน เป็นต้น
2. ภาพที่มีวัตถุเกิดปรากฏขึ้น สำหรับภาพที่ไม่มีความหมายหรือเป็นภาพที่มีความหมายกำกวม หรือภาพที่สามารถแปลได้หลายความหมาย หรือภาพที่โฟกัสระยะใกล้ จะไม่นำภาพนั้นเข้ามาวิเคราะห์สำหรับการหาความหมายภาพข้อมูลที่น่ามาวิเคราะห์เป็นข้อมูลภาพดิจิทัลที่เห็นวัตถุชัดเจน
3. ข้อมูลหรือวัตถุนภาพที่เป็นอินพุตบนภาพจะถูกแก้ไข เป็นคำศัพท์และขนาดมาก่อนด้วยกระบวนการ รู้จำของการประมวลผลภาพในระดับต่ำ (low-level image)
4. ในการสร้างแบบจำลองภาพ (image representation) ข้อมูลจะถูกจัดเก็บมาในรูปแบบของฟีเจอร์ข้อมูลไว้แล้ว
5. ภาพจากคลังภาพจะถูกตั้งค่าเริ่มต้นโดยการใส่ข้อมูลวัตถุนภาพพร้อมความสัมพันธ์ภาพไว้ก่อนแล้ว
6. ภาพที่ถูกลำมาใช้เป็นข้อมูลทดลองจากคลังภาพที่เหมาะสมกล่าวรายละเอียดในส่วนถัดไป
7. การเก็บผลข้อมูลเบื้องต้นของความหมายของท่าทางการเคลื่อนไหวจากกลุ่มนักศึกษาและบุคคลทั่วไป เป็นกลุ่มบุคคลที่ตัดสิน ความหมายของท่าทาง สำหรับการทดลองในคลังภาพที่กล่าวมา

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ผลต่อสังคม

1. เพื่อนำเอากระบวนการนี้มาประยุกต์ใช้กับการค้นหาภาพบนห้องสมุดดิจิทัลได้
2. เพื่อได้แนวทางการแปลและการตีความหมายภาพ
3. ช่วยให้การใช้คำศัพท์ในการค้นหาข้อมูลภาพได้ผลตามความหมายของภาพมากขึ้น

1.5.2 ผลต่อมหาวิทยาลัย

1. สร้างกลุ่มนักวิจัยที่เป็นลักษณะของสาขาทางการประมวลผลภาพในสาขาย่อย Semantic Image Processing
2. เพื่อสนับสนุนและเสริมสร้างความรู้ความสามารถของการพัฒนาตนเอง และ กลุ่มงานในสาขาเทคโนโลยีสารสนเทศ
3. เพื่อเพิ่มตัวชี้วัดให้กับองค์กร สถาบันในส่วนของงานวิจัยและพัฒนา
4. สร้างชื่อเสียงให้กับมหาวิทยาลัย เมื่อมีบทความวิจัยลงในวารสารต่างประเทศ บทความวิจัยที่นำเสนอในการประชุมระดับชาติ และนานาชาติ

1.5.3 ผลต่อกลุ่มผู้วิจัย

1. พบแนวทางในการคิดค้นสิ่งใหม่ๆ ในการจัดกลุ่มและกระบวนการแปลความหมายของภาพ
2. สามารถนำผลวิจัยมาเขียนบทความลงวารสารนานาชาติ และร่วมประชุมวิชาการระดับชาติและนานาชาติ
3. พัฒนาความรู้ใหม่ๆ หลักการแนวคิดใหม่ ให้เกิดขึ้นในกลุ่มของนักวิจัย

1.6 นิยามคำศัพท์

1. การประมวลผลภาพ (Image processing) หมายถึง การนำภาพมาผ่านกระบวนการเพื่อประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง (ภาพถ่าย) หรือภาพวีดิทัศน์ (วิดีโอ) และนำมาใช้งาน
2. การจำแนกประเภทข้อมูลภาพ (Image Classification) หมายถึง การแยกแยะภาพที่มีคุณลักษณะเดียวกันออกเป็นกลุ่ม ๆ
3. การจำแนกความหมายของภาพ (Semantic Image Classification) หมายถึง การนำภาพมาแบ่งเป็นกลุ่มโดยพิจารณาจากความหมายโดยรวมของภาพ

4. การค้นคืนข้อมูล (Image Retrieval) หมายถึง การค้นหาข้อมูลของภาพจากระบบฐานข้อมูล
5. วัตถุ (Object) หมายถึง ส่วนของวัตถุบนภาพ ยกตัวอย่างเช่น คน, ต้นไม้ เป็นต้น
6. คำหลัก (Keyword) หมายถึง คำที่มีความหมายได้ใจความใช้แทนวัตถุบนภาพ
7. แท็ก (Tag) หมายถึง การกำหนดคำ หรือ คำศัพท์บนภาพ หรือเรียกว่าการ labeled
8. คุณลักษณะข้อมูล หรือ ฟีเจอร์ (Feature) หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป ในงานวิจัยนี้ข้อมูลที่สกัดออกมาเป็นมุมมองเสาที่ถูกกระทำกับแกนหมุน กล่าวรายละเอียดในส่วนถัดไป
9. การใช้พลังงาน (Energy expenditure) หมายถึงพลังงานที่ได้จากการคำนวณจากท่าทางกายภาพของร่างกายมนุษย์

บทที่ 2

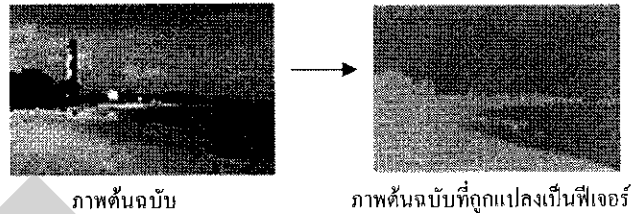
ทฤษฎีและงานวิจัย

2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาคือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร ซิดจำกัดนี้จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวมาข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้โดยสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ ยิ่งมีความจำเป็นอย่างยิ่ง

สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ ยกตัวอย่างเช่น สี (color) ลวดลาย (texture) รูปร่าง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) [Th. Gevers 2000][F. Porikli 2005] กระบวนการนี้สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานที่นำไปสืบค้น โดยภาพผลลัพธ์ส่วนใหญ่มักจะเป็นภาพที่มีคุณลักษณะไม่ซับซ้อนมากนัก เช่น โทนสี หรือ รูปร่างที่แตกต่างกันอย่างเด่นชัด ดังภาพที่ 2.1 แสดงการค้นหาของ SIMPLIcity [Jia Li,2003][James Z. Wang,2001] ด้วยคุณลักษณะพีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างสิ้นเชิงที่อยู่ในหมวดหมู่เดียวกัน คือ ภาพชายหาด หรือ ชายทะเล แต่มีคุณลักษณะของโทนสี และลวดลาย ที่แตกต่างกันอย่างเห็นได้

ชัดเจน เมื่อนำมาจำแนกด้วยการประมวลผลภาพระดับต่ำ (low-level feature) แล้วนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน



ก. ภาพต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นฟีเจอร์ระดับต่ำ

SIMPLICITY
Semantics-sensitive Integrated Matching for Picture Libraries

Option 1 -> Image ID or URL: Option 2 -> **Random** Option 3 -> Click an image to find similar images

ข. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ ก.

ภาพที่ 2.1 การประมวลผลแบบค้นคืนด้วยคุณลักษณะฟีเจอร์ระดับต่ำ¹

แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามปรับปรุงแปลงอัลกอริทึมด้วยการประมวลผลภาพระดับต่ำ เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [M. Flickner, 1995][W. Ma, 1997] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการค้นคืนภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่น การรวมเทคนิคด้วยคุณลักษณะสีและรูปร่างของภาพ เพื่อทำการค้นคืนภาพ หรือมีการใช้อัลกอริทึมเพื่อทำการสกัดข้อมูลภาพเป็นฟีเจอร์ และนำฟีเจอร์นำมาใช้ในการ ค้นคืนภาพในรูปแบบที่แตกต่างกันออกได้ แต่ในความเป็นจริงแล้วนั้นลักษณะการ

¹ http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi

มองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือมองจากชนิดของวัตถุของภาพ ดังแสดงในภาพที่ 2.2 แสดงภาพชายทะเล ภายในภาพประกอบด้วยวัตถุหลายชนิดได้แก่ ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วยวัตถุอื่นๆ มนุษย์ สุนัข หรืออาคาร เป็นภาพ ประเภทเดียวกันหรือภาพที่สื่อความหมายอย่างเดียวกัน โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือ รูปทรงแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการค้นคืนที่ใช้คุณลักษณะพีเจอร์ ระดับต่ำเพียงอย่างเดียว อาจจะได้ผลลัพธ์ที่ไม่ตรงกับความต้องการอย่างแท้จริง



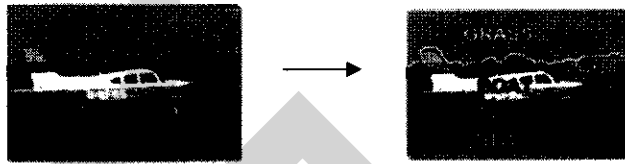
ภาพที่ 2.2 ตัวอย่างขององค์ประกอบของภาพ²

การวิจัยในส่วนแรกเป็นความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดข้อมูล ภายในภาพ (feature extraction) ออกมาเป็นส่วน ๆ แล้วใช้ข้อมูลระดับต่ำเพียงแบบเดียว เพื่อทำการ สืบค้นข้อมูลภาพเท่านั้น วิธีการค้นคืนด้วยข้อมูลระดับต่ำเพียงแบบเดียวแล้วนั้นจะทำให้ได้ผลลัพธ์ ของภาพที่มีความหมายไม่ครบถ้วน ทำให้ต้องมีการปรับปรุงพัฒนาอย่างต่อเนื่อง ด้วยเหตุนี้เองทำให้ มีกลุ่มนักวิจัยที่สนใจในเรื่องของ การจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการ จัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความ ต้องการมากที่สุด

แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของ ภาพแทน การสืบค้นแบบข้างต้น เรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B,2002] [Galleguillos C.,2010][Galleguillos C., 2011] และแทนวัตถุอื่นๆ ด้วย

² <http://www.corbisimages.com/>

คำหลัก (keyword) บนภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อ วัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass”, “plant”, “boat”, “sky” เป็นต้น ดังแสดงในภาพที่ 2.3 [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมายของคำศัพท์ที่มีสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำหลัก (synonym) [Zhao T., 2001] [Benitez, A.B., 2002] [Kobus B., 2001] [Philippe M., 2002] เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น จะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด

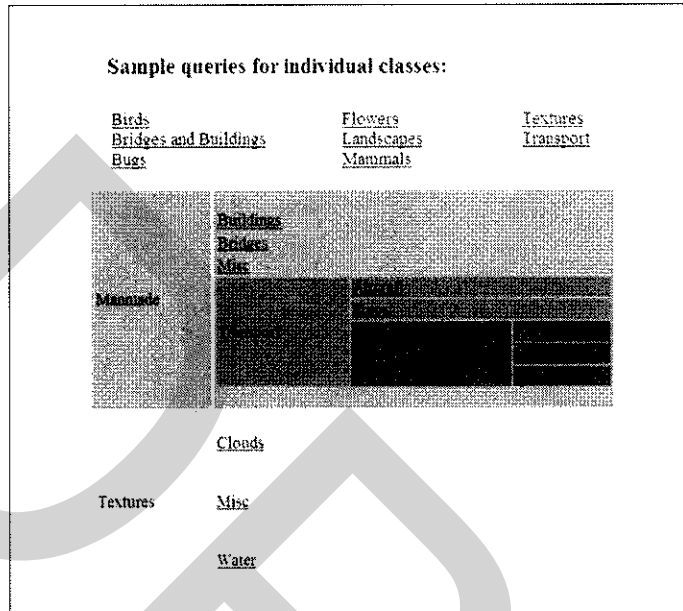


ภาพที่ 2.3 ตัวอย่างการแท็กองค์ประกอบของภาพ³

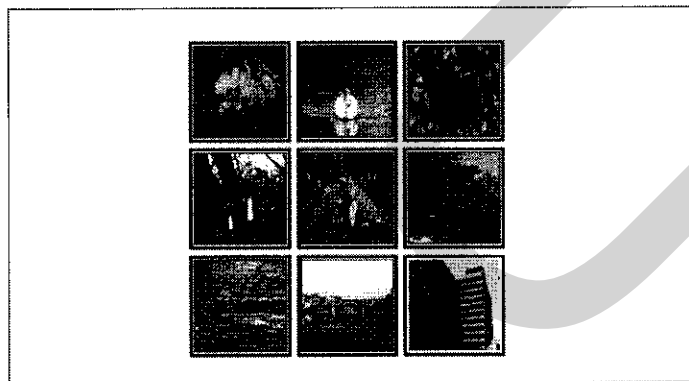
นักวิจัยบางกลุ่มใช้เทคนิคเพื่อสร้างความสัมพันธ์ (relationship) ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001] [R. Zhao, 2002] [Philippe M., 2002] ด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee L., 2003] เช่น การเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch” “on” “top” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความ (context) เพื่อบรรยายความหมายของภาพ [Mathias Lux, 2003] [Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพนั้นๆ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” แต่ภาพที่นำมาใช้จัดเก็บนั้นมักจะเป็นจะภาพส่วนตัว (personal images) ทำให้การค้นคืนจำกัดเพราะคำบรรยายส่วนใหญ่จะเป็นคำเฉพาะเจาะจง จึงไม่นิยมเท่าที่ควร บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่างๆจนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ดังนั้นในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่ นอกเหนือ หรือเกินความจำเป็นโดยใช้เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้น ข้อมูล อาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย ทำให้การค้นคืนยังเกิดความสับสนปนเป่กันของภาพผลลัพธ์ที่ได้มา ดังแสดงในภาพที่ 2.4 ก. [Q. Iqbal, 2002] แสดงตัวอย่างการค้นคืนภาพที่มีการจัดเป็นหัวข้อ (topic) จากภาพที่ 2.4 ข. แสดงภาพต้นฉบับเพื่อ

³ <http://vision.ucsd.edu/project/context-based-object-categorization>

สร้างคุณลักษณะหรือฟีเจอร์ที่คล้ายกันเพื่อใช้ในการค้นคืนในที่เก็บข้อมูลภาพ จากภาพที่ 2.4 ค. แสดงภาพต้นฉบับมีคำหลักเป็น “bridges” และกลุ่มผลลัพธ์ที่ได้หลังจากการใช้ฟีเจอร์ในการค้นคืนภาพ แต่ผลลัพธ์ที่ได้ออกมานั้นจะมีการผสมกับภาพอื่น ๆ ที่ไม่ตรงกับคำหลักด้วย



ก. การค้นคืนด้วยการเลือกคำหลักที่เป็นหมวดหมู่



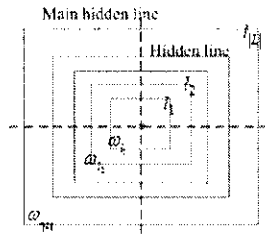
ข. การค้นคืนด้วยการเลือกภาพต้นฉบับ



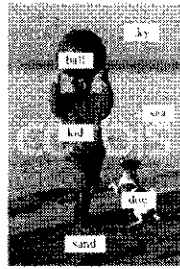
ค. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ
ภาพที่ 2.4 การแก้กองค้ประกอบของภาพ⁴

การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแก้ไว้บนภาพยิ่งมีการแก้ ข้อมูลบนภาพมากยิ่งขึ้นสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแก้ข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ สำหรับการวิเคราะห์ความหมายของภาพจะได้รับการรับรู้ของมนุษย์ที่มองภาพนั้น จะมีกฎเกณฑ์ของการรับรู้จากการมองภาพได้กล่าวไว้เป็นทฤษฎี โครงสร้างสเกตตรีตรอน ดังแสดงในภาพที่ 2.5 ก. [Rudolph A., 1974] ได้มีการนำมาประยุกต์ใช้วิเคราะห์ในส่วนของเฉพาะการการจำแนกความหมายของภาพโดยรวม เพื่อให้ได้ความหมายของภาพอย่างแท้จริง โดยใช้ทฤษฎีการมองของมนุษย์ที่มีการพิจารณาจาก ตำแหน่งและขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพ ด้วยการวัดโครงสร้างกับภาพต้นฉบับที่มีการแก้ ดังแสดงในภาพที่ 2.5 ข. และ ค. [N. Chinpanthana., 2010] [นิตยสาร 2552] แต่อย่างไรก็ตามการคำตำแหน่งในส่วนของภาพนั้นยังไม่เพียงพอต่อการนำมาใช้ สำหรับการแปลความหมายภาพ

⁴ http://amazon.ece.utexas.edu/~qasim/sample_queries.htm



ก. โครงสร้างสเกตตรีตรอน



ข. ภาพแท็กด้วยคำหลัก



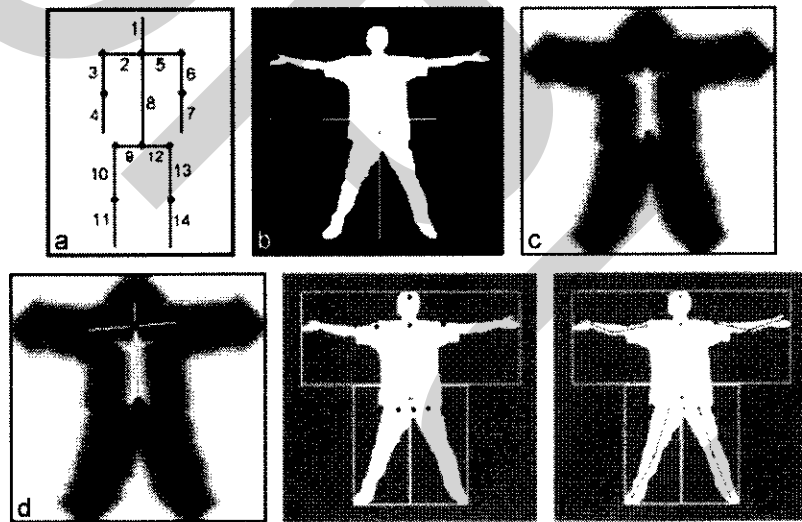
ค. แม็พด้วยโครงสร้างสเกตตรีตรอน

ภาพที่ 2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน

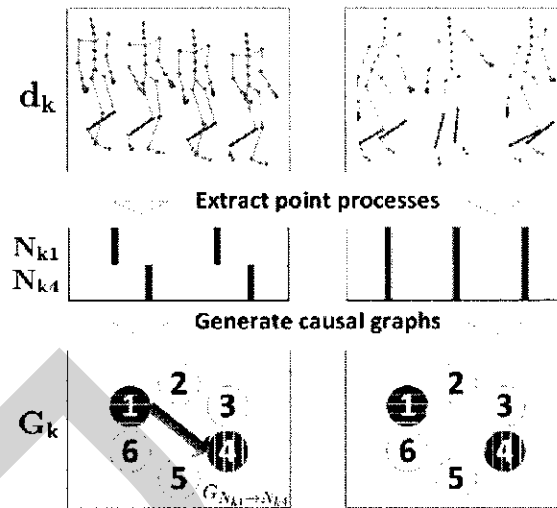
ยังมีวิธีการที่ผสมผสานระหว่างการใช้แท็กคำหลักบนภาพกับการสร้างความสัมพันธ์ระหว่างวัตถุด้วยมัลติเพล็กซ์ (multiple kernel) ของพื้นที่วัตถุบนภาพ [Galleguillos C., 2011] [Galleguillos C., 2010] ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเสียส่วนใหญ่ และมีนักวิจัยบางกลุ่มพยายามที่แก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น จึงได้พัฒนาซอฟต์แวร์เพื่อให้เกิดความสะดวกสบายในการใส่ข้อมูล และการใส่ข้อมูลที่มีแบบแผนแน่นอน ยกตัวอย่างเช่น Caliph & Emir [Mathias Lux, 2009], Annosearch [Xin-Jing, 2008], CAMEL [Apostol Paul N., 2001] เป็นต้น แต่อย่างไรซอฟต์แวร์ส่วนใหญ่ นั้น จะทำการสืบค้นภาพในรูปแบบของการเทียบคำหลักเป็นคำต่อคำตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นคำตอบที่ได้จะไม่ได้ขึ้นกับความหมายของภาพ แต่ขึ้นกับคำศัพท์ที่มีการเก็บข้อมูลลงไปเท่านั้น ความหมายของภาพยังไม่ได้้อย่างแท้จริง เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น จึงทำให้ความหมายของภาพโดยรวมนั้นยังไม่ได้สื่อออกมาให้เห็นอย่างเด่นชัด ซึ่งในความเป็นจริงแล้วความหมายของภาพ โดยเฉพาะภาพส่วนบุคคล (Personal image)

หลายงานวิจัยในสาขาคอมพิวเตอร์วิชัน (Computer vision) ที่พยายามแปลความหมายของการแสดงออกของท่าทางของมนุษย์ (Human action) [Weilong Yang 2010][Weilong Yang 2009] เพื่อบอกถึงกิจกรรม (Activity)ที่กำลังกระทำบนภาพและสามารถนำการตีความหมายนั้นมาใช้บนแอปพลิเคชันต่างได้ เช่น Surveillance, Entertainment, Human-computer interaction, Image หรือ Video search เป็นต้น ปัจจุบันการรู้จำท่าทางมนุษย์ (Human action recognition) บนภาพนิ่ง (Still image) ยังเป็นสิ่งที่จำเป็นและเป็นพื้นฐานของการใช้งานและดัดแปลงบนภาพวิดีโอ หลายหน่วยงานมีการพัฒนาวิธีการอย่างต่อเนื่องเกี่ยวกับการเข้าใจการกระทำบนการกระทำของมนุษย์ที่แสดงออกบนภาพ Wang et al., [Y.Wang 2006] ใช้วิธีการหาความแตกต่างของคัสเตอร์ของท่าทางมนุษย์ (Cluster different human poses) จากการ Deformable shape matching ใน

งานวิจัยของ Thureau and Hlaváč [C. Thureau 2008] จะใช้การแทนรูปแบบของท่าทางด้วยฮีสโตแกรมของท่าทางหลักโดยใช้ Nonnegative matrix factorization Albu et al., [A. Branzan Albu 2008] ได้พยายามสกัดพีเจอร์ที่ได้จากท่าทางมนุษย์บนภาพเงา (Silhouettes) แบบ 1 มิติที่มีการแทนค่าเป็นโปรเจกชันของร่างกายมนุษย์จากส่วนต่างๆที่มีการกำหนดไว้ถึง 6 ส่วนแบ่งซึ่งจะรวมทั้ง ร่างกาย แขนและขา ดังแสดงในภาพที่ 2.6 สำหรับการสกัดพีเจอร์ที่ได้จากภาพเงาในบางครั้ง การทับกันของส่วนร่างกายอาจจะทำให้การสกัดพีเจอร์ได้ไม่ครบ แต่อย่างไรก็ตาม Saehoon Yi [Saehoon Yi 2012] ได้พยายามแยกแยะท่าทางของมนุษย์ ด้วยวิธี Sparse Granger Causality Graph Model เป็นการแทนตำแหน่งร่างกายด้วยโหนดบนกราฟ แต่ละเส้นเอ็ดจ์บนกราฟมีการแทนค่าด้วยตัวถ่วงน้ำหนักเพื่อแทนการเคลื่อนไหวของส่วนของร่างกาย ดังแสดงในภาพที่ 2.7 แสดงตัวอย่างของ ท่าทางการเดิน และท่าทางการกระโดด ซึ่งมีค่าความแตกต่างรูปแบบในแต่ละท่าทางกราฟที่แสดงเป็น Granger causality graph ใช้เก็บค่าคุณลักษณะของท่าทางเพื่อทำการแยกแยะท่าทางต่อไป



ภาพที่ 2.6 ตัวอย่างการแบ่งส่วนบนร่างกายด้วยภาพเงาและการหามุมจากภาพเงา
[A. Branzan Albu 2008]



ภาพที่ 2.7 ตัวอย่างท่าทางของมนุษย์เพื่อจัดเก็บในรูปแบบ Granger causality graph
[Sachoon Yi 2012]

จากงานวิจัยที่มีการนำเสนอในรูปแบบของการสกัดข้อมูลท่าทางมนุษย์นั้นยังไม่สามารถที่จะใช้ในการแทนความหมายได้อย่างแท้จริง ดังนั้นในงานวิจัยนี้จึงนำเสนอรูปแบบใหม่ของการรู้จำท่าทางมนุษย์ โดยนำหลักการของการใช้พลังงาน (energy expenditure) การเคลื่อนไหวของท่าทางร่างกาย (human action) เข้ามาวิเคราะห์ท่าทางมนุษย์ เพื่อสามารถนำไปใช้กับการแปลความหมายภาพ และการสืบค้นข้อมูลภาพได้อย่างมีประสิทธิภาพมากที่สุด จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอการปรับปรุงวิธีการรู้จำท่าทางมนุษย์ (human action recognition) เพื่อให้ได้ผลลัพธ์ของพฤติกรรมมนุษย์ที่อยู่บนภาพอย่างแท้จริง โดยใช้วิธีการสกัดข้อมูลจากท่าทางการเคลื่อนไหวของร่างกายมนุษย์ (Human action movement) ด้วยโมเดลการใช้พลังงาน (Energy Expenditure) และทำเปรียบเทียบการรู้จำท่าทางทั้งหมด 3 วิธีการ [R. O. Duda 1973] [T. Mitchell 1997] โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network), แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

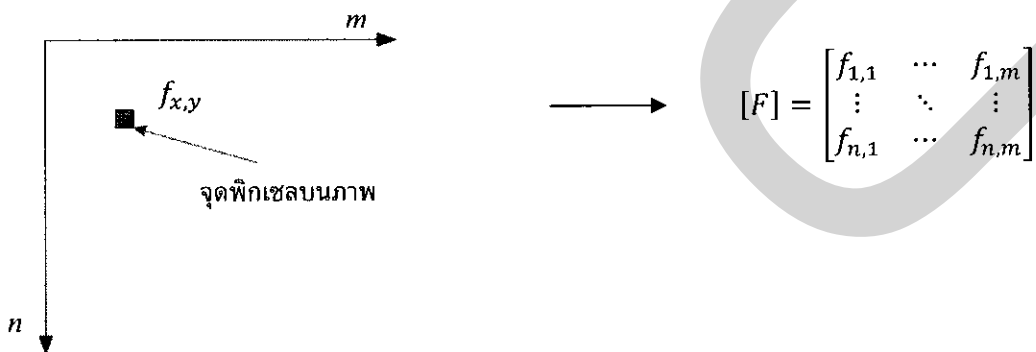
ในงานวิจัยนี้ได้นำทฤษฎีมาใช้ในการพิจารณาดังนี้

- การประมวลผลภาพดิจิทัล
- การแยกแยะความหมายภาพ

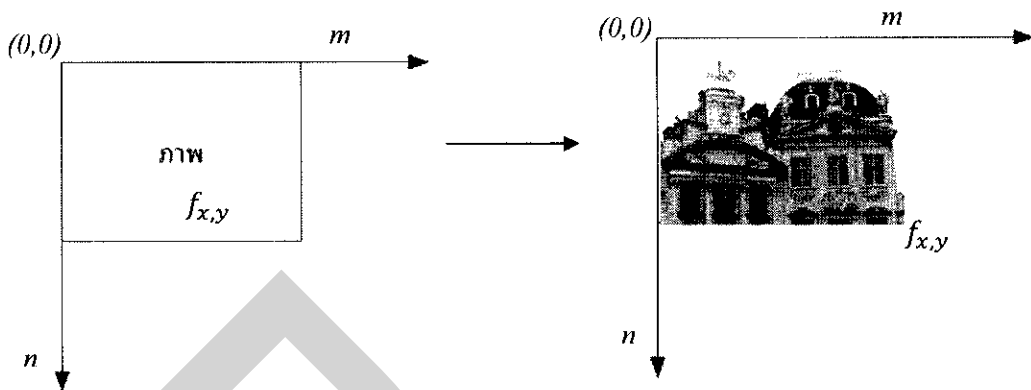
- การวัดประสิทธิภาพ

2.2 การประมวลผลภาพดิจิทัล

ภาพจะถูกจัดเก็บเป็นสัญญาณดิจิทัล (digital) [นิตยสาร 2553] [R.C. Gonzalez 2002][R. C. Gonzalez 2004] และทำการจัดเก็บข้อมูลภาพในหน่วยความจำ ในรูปของตัวแปรแบบอะเรย์ (array) โดยค่าในแต่ละช่องของอะเรย์แสดงถึงค่าความเข้มของแสง (intensity) ของแต่ละจุดของแต่ละจุดของภาพที่เรียกว่าพิกเซล (pixel) และตำแหน่งของช่องอะเรย์เป็นตัวกำหนดตำแหน่งของจุดภาพดังแสดงในภาพที่ 2.4 เมื่อนำภาพมาแสดงในรูปของ 2 มิติ ตามแนวแกน x และ y สามารถเขียนในรูปของฟังก์ชันของภาพสามารถแสดงในรูปของ $i = f(x, y)$ ในรูปของ 2 มิติ โดย i หมายถึง ความสว่างหรือความเข้มของแสง และ (x, y) เป็นพิกัดในแบบจำลองของภาพ (image model) ค่าของฟังก์ชัน $f(x, y)$ ดังกล่าวเป็นค่าความเข้ม ซึ่งเปลี่ยนไปตามตำแหน่งหรือพิกัดแบบจำลองภาพ ค่าความเข้มของแสงนี้เป็นค่าดิจิทัลหรือเชิงตัวเลข ดังนั้น เมื่อแสดงเป็นตัวเลขของค่าความเข้มในแต่ละพิกเซลเรียงกันทั้งภาพจะเห็นว่าเป็นเมทริกซ์ สิ่งนี้แบบจำลองต่างจากแบบจำลองทางคณิตศาสตร์ คือ จุดกำเนิด (Coordinate Origin) ที่ตำแหน่ง $(0,0)$ ซึ่งอยู่ที่จุดซ้ายบนสุดของสเกลต่างจากจุดกำเนิดของแบบจำลองทางคณิตศาสตร์ที่อยู่จุดซ้ายล่างสุด ดังนั้นการนำเอาแบบจำลองทางคณิตศาสตร์มาประยุกต์ในการประมวลผลภาพ จำเป็นต้องเปลี่ยนจุดกำเนิดให้ถูกต้องด้วย ดังแสดงในภาพที่ 2.8



ภาพที่ 2.8 การแสดงตำแหน่งของพิกเซลบนภาพดิจิทัลด้วยเมทริกซ์

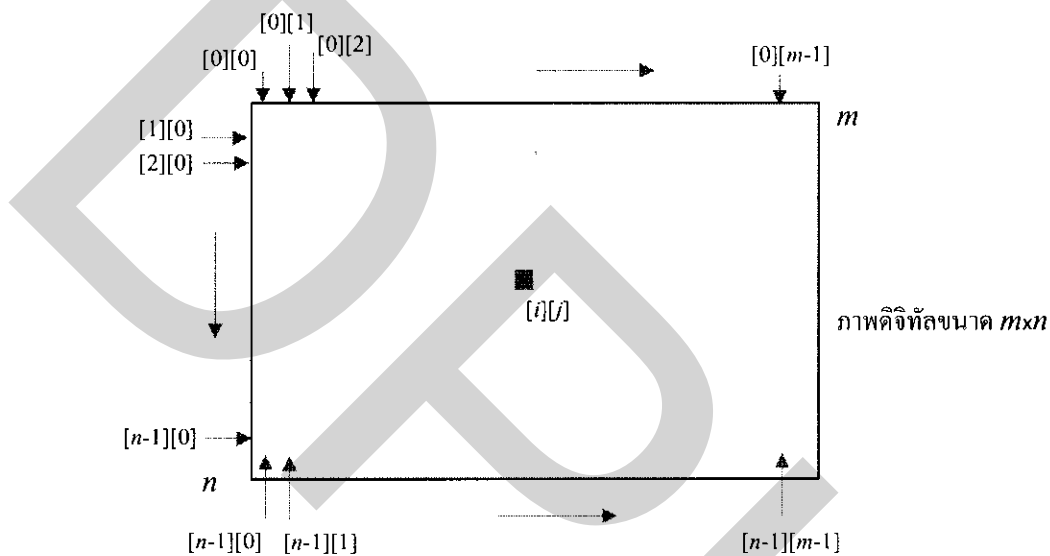


ภาพที่ 2.9 แบบจำลองของภาพด้วยระนาบ

ในหน่วยความจำ จะทำการจัดเนื้อที่ในการเก็บภาพ สามารถคำนวณได้จาก $m \times n \times b$ เมื่อ b เป็นจำนวนเต็มที่แทนจำนวนบิตของข้อมูลในแต่ละจุดภาพ ตัวอย่างถ้า b มีค่าเท่ากับ 8 บิต จะสามารถเก็บความแตกต่างของระดับสีที่เป็นไปสูงสุด 256 ระดับ ค่า m และ n จะเป็นตัวบอกถึงความละเอียดของภาพ สำหรับคอมพิวเตอร์ทั่วไปในระบบ VGA (Video Graphic Array) จะมีขนาด 640×480 , 800×600 และ 1024×768 จุด เป็นต้น การกำหนดความละเอียดจะขึ้นอยู่กับงานที่จะใช้ ในงานบางอย่างใช้ความละเอียดเพียง 30×50 จุด ก็พอซึ่งความละเอียดนั้นจะขึ้นกับงานที่จะใช้ ในบางงานจะใช้ความละเอียดถึง 1000×1000 จุด ก็ยังไม่พอ จากภาพที่ 2.8 สมมติให้ภาพแทนเป็นตัวแปรชื่อ x เป็นตัวแปรแบบอะเรย์ขนาด $m \times n$ (m แทน แถว และ n แทน คอลัมน์) ที่ใช้เก็บภาพขนาด $m \times n$ จุด และ ค่าของความเข้มของแสง (ค่าความสว่าง) ของจุดภาพในแถวที่ 5 คอลัมน์ที่ 4 จะตรงกับค่าของข้อมูล x เป็นรูปของ (5,4) จะเห็นว่าใช้ตำแหน่งของจุดภาพทั้งสองแกนเป็นตัวชี้ค่าข้อมูลในอะเรย์ ดังแสดงการเรียงตัวของข้อมูลบนอะเรย์ในภาพที่ 2.10 และการกำหนดความละเอียดของภาพ (image resolution) จากการกำหนดขนาดของพิกเซลตัวอย่างเช่น 1 ไมครอนต่อพิกเซล ($\mu\text{m}/\text{pix}$) 1 มิลลิเมตรต่อพิกเซล (mm/pix) เป็นต้น ในงานที่ต้องการทราบตำแหน่งหรือขนาดของวัตถุที่วัดเป็นค่าจริง เราสามารถที่จะคำนวณได้จาก

$$\text{Resolution} = \frac{\text{Field of vision in Y direction (mm)}}{\text{Number of pixels in Y direction}}$$

โดยปกติแล้วในการเก็บข้อมูลภาพโดยเครื่องมือต่าง ๆ จะเก็บตามมาตรฐานของโทรทัศน์ซึ่งมีอัตราส่วน x ต่อ y เท่ากับ 4:3 สำหรับเครื่องมือเก็บข้อมูลภาพที่ไม่เป็นไปตามอัตราส่วน 4:3 เมื่อนำภาพนี้ไปแสดงในจอภาพมาตรฐาน จะทำให้ภาพที่แสดงนั้นมีขนาดของจุดภาพไม่เป็นสี่เหลี่ยมจัตุรัส เช่น ในบางระบบอาจจะใช้ความละเอียดในการแสดง เท่ากับ 640×580 ซึ่งจะทำให้ขนาดของจุดภาพที่ได้มีขนาดของด้านกว้างมีความยาวมากกว่าด้านสูง เมื่อมีการกำหนดให้ขนาดของบิตต่อจุดมากขึ้น จะทำให้จำนวนของสีมากขึ้นด้วย ตัวอย่างเช่น 1 บิต = 21 จะได้ 4 สี 2 บิต = 22 จะได้ 4 สี 4 บิต = 24 จะได้ 16 สี 8 บิต = 28 จะได้ 256 สี 16 บิต = 216 จะได้ 65536 สี เป็นต้น



ภาพที่ 2.10 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์

จากคุณลักษณะของภาพดิจิทัลที่กล่าวมาข้างต้นนั้น เป็นการเก็บข้อมูลภาพเป็นแบบ เมทริกซ์ ที่มีถึง 3 ระนาบ (dimension) การประมวลผลภาพดิจิทัล (digital image processing) เป็นการเรียกใช้ขั้นตอนหรือกระบวนการที่มากกระทำบนภาพ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของภาพ ให้ได้ภาพใหม่ที่มีคุณสมบัติตามวัตถุประสงค์ที่นำไปใช้งาน เช่น การปรับให้ภาพมีความคมชัดมากขึ้น (enhancement) หรือการบีบอัดข้อมูลภาพ (compression) เพื่อประหยัดพื้นที่ในการจัดเก็บข้อมูล หรือการฝังลายน้ำ (watermark) เพื่อป้องกันการลักลอบการใช้ภาพที่ไม่ได้รับอนุญาต เป็นต้น สำหรับการประมวลผลภาพระดับสูงด้วยคอมพิวเตอร์ สามารถทำได้โดย นำภาพที่ได้มาจากกล้องหรือ image source ต่าง ๆ ซึ่งเป็นสัญญาณอนาล็อก แล้วนำมาแปลงเป็นสัญญาณดิจิทัลที่มีลักษณะเป็นรหัสเชิงตัวเลขฐานสอง (binary) ประกอบด้วยตัวเลข 0 และ 1 ที่สามารถใช้รูปแบบทางคณิตศาสตร์

เข้ามาช่วยในการคำนวณและการประมวลผลข้อมูลภาพ การดำเนินงานวิจัยสำหรับการจัดกลุ่มความหมายของภาพได้มีผู้วิจัยจำนวนมากที่ศึกษาและทำการทดลองโดยทั่วไป แบ่งเป็น 2 ระดับ [A. Gupta, 1997] ดังนี้

2.2.1 การประมวลผลภาพเบื้องต้น

สำหรับขั้นตอนการประมวลผลภาพเบื้องต้น (image preprocessing) ในรูปแบบของการหาบริเวณที่ต้องการในรูปแบบอัติโนมติ ยังคงเป็นงานวิจัยที่ยังหาข้อยุติไม่ได้ โดยเฉพาะที่เป็นลักษณะของระบบเรียลไทม์ (real time) ด้วยแล้วนั้นจะต้องคำนึงถึงเวลาในการคำนวณของอัลกอริทึม (computational cost of algorithm) ที่เป็นสิ่งที่จำเป็นค่อนข้างมาก สำหรับระบบเรียลไทม์ ที่มีการแปรเปลี่ยนรูปแบบของพื้นหลัง (modeling of background) จะทำให้เกิดกระบวนการในการสกัดวัตถุที่ต้องการขึ้นมา ซึ่งเทคนิคที่เรียกว่า Gaussians Mixture Model [C. Stauffer, 2000] เป็นเทคนิคที่ค่อนข้างประสบความสำเร็จมากแต่อย่างไรก็ตามยังคงต้องมีส่วนภาพพื้นหลังที่ไม่มีการเปลี่ยนแปลงมากนัก เพราะฉะนั้นปัญหาที่เกิดขึ้นยังคงเกิดขึ้นเมื่อพื้นหลังมีการแปรเปลี่ยนและยังคงเป็นปัญหาที่ยังคงต้องมีการแก้ไข ฉะนั้นในงานวิจัยนี้จึงข้ามในส่วนการวิเคราะห์การแปรเปลี่ยนของพื้นหลังที่ไม่คงที่ และได้มีการนำเทคนิคของฮิสโตแกรม (Histogram matching) [F. Porikli, 2005] [D. Comaniciu, 2003] [D. Comaniciu, 2000] ที่ไม่ได้มีการนำส่วนของพื้นหลังเข้ามาเกี่ยวข้อง เพื่อทำการเปรียบเทียบแต่อย่างไรก็ตาม ด้วยคุณลักษณะของเทคนิคฮิสโตแกรมไม่สามารถที่จะรองรับตำแหน่งของพิกเซล (pixel location) ทำให้ยังคงมีปัญหาเมื่อมีวัตถุที่ทับซ้อนกัน จนไม่สามารถที่จะประมวลผลได้อย่างถูกต้อง เทคนิคสหสัมพันธ์ (Correlation) [A.J. Lipton, 1998] [S. Wong, 2005] จะมีการทำงานที่ไม่สูญเสียในข้อมูลส่วนของสเปเชียลทำให้สามารถทดแทนวิธีการของฮิสโตแกรมได้ และได้มีการใช้เทคนิคสหสัมพันธ์เพื่อทำการหาขอบของวัตถุซึ่งวิธีการนี้ถูกเรียกว่า Edge-Enhanced Normalized Correlation (EENC) [Javed Ahmed, 2008] เพื่อทำการแก้ไขปัญหาต่างๆที่เกิดจากการหาวัตถุที่เกิดจากการทับซ้อนการ หรือมีปัญหาจากสัญญาณรบกวน หรือวัตถุที่มีการเปลี่ยนแปลงทิศทางจากการหมุน เป็นต้น และการใช้เทคนิค EENC สามารถทำงานได้ดีสำหรับการแมชชีงส่วนของพื้นที่ (matching region) พร้อมทั้งยังเป็นเทมเพลต (template) ที่ใช้ทำงานได้อย่างรวดเร็ว

2.2.2 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้องค์ความรู้หรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมที่ทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไป เนื่องจากยังไม่มีความสำเร็จเท่าที่ควรและพยายามที่เข้าใช้การประมวลผลภาพระดับสูงอย่างเดียว แต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือ ฟีเจอร์ต่างๆเข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification)

2.3 การจำแนกประเภทข้อมูลภาพ

การจำแนกประเภทข้อมูลภาพ (image classification) เป็นการนำเอาข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ มาผ่านกระบวนการขั้นตอนการแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูล หรือ ฟีเจอร์ ที่เก็บรวบรวมมาได้ รวมทั้งกระบวนการ หรือวิธีการที่ใช้แยกแยะข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการแยกประเภทข้อมูลของภาพเพื่อเปรียบเทียบกันทั้งหมด 4 วิธี

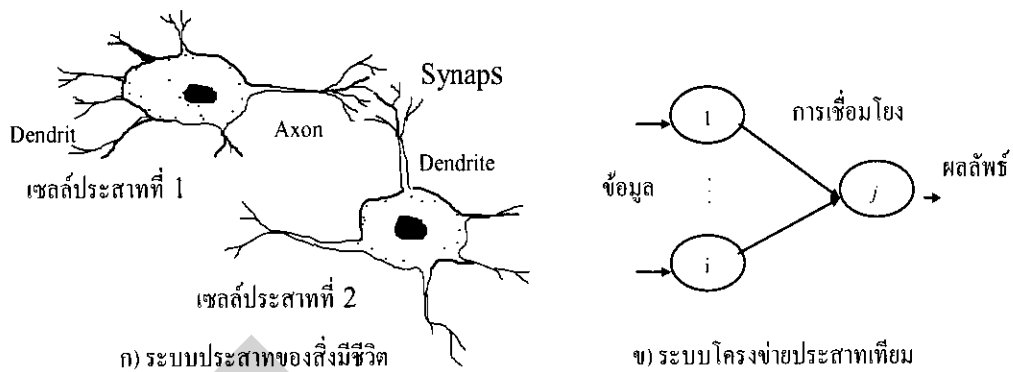
- ทฤษฎีเครือข่ายแบบเบย์ (Bayesian Network)
- โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network)
- ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-organizing maps)
- ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines)

2.3.1 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (artificial neural network) [R. O. Duda, 1973] [Lippmann, R.P., 1987] หรือเรียกว่า ข่ายงานประสาท (neural network หรือ neural net) เป็นโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำแบบรูป (pattern recognition) และการอุปมาความรูป (knowledge deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ ข่ายงานประสาทเทียมถูกพัฒนามาจากวัตถุประสงค์ดังนี้

- การเลียนแบบการทำงานของเซลล์สมองของมนุษย์
- การหาความสัมพันธ์ของเหตุและผลซึ่งยังหาฟังก์ชันออกมาโดยตรงไม่ได้
- การทำให้คอมพิวเตอร์คิดและตัดสินใจได้ด้วยตัวเอง

แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ประกอบด้วย เซลล์ประสาท หรือ “นิวตรอน” (neurons) และ จุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรท์” (dendrite) เป็นข้อมูลเข้า (input) และปลายในการส่งกระแสประสาทเรียกว่า “แอกซอน” (axon) เป็น ผลลัพธ์ (output) ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่นๆ ต่อไปผ่านทางแอกซอน ดังแสดงในภาพที่ 2.11 จากรูป เซลล์ร่างกาย (cell body) คือ ตัวของเซลล์ประสาท โดยสัญญาณไฟฟ้าจะเข้ามาทาง เดนไดรท์ซึ่งจะมีลักษณะเป็นแขนงย่อยและมีจำนวนมาก จากนั้น เซลล์ร่างกายจะทำหน้าที่ประมวลผลสัญญาณ และยิงสัญญาณออกไปทางเดนไดรท์ให้เซลล์ประสาทอื่นต่อไป



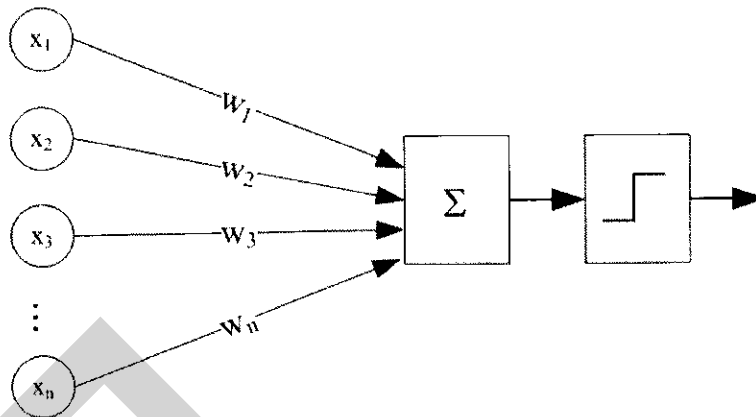
ภาพที่ 2.11 เซลล์ระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม

▪ หลักการทำงานข่ายงานประสาท

สำหรับในคอมพิวเตอร์นิวรอน [Chester M., 1993][Smith M., 1993] ประกอบด้วย ข้อมูลเข้า และ ข้อมูลออกเหมือนกัน โดยจำลองให้ ข้อมูลเข้า แต่ละอันมีค่าน้ำหนัก (weight) เป็นตัวกำหนด น้ำหนักของ ข้อมูลเข้า โดย นิวรอนแต่ละหน่วยจะมีค่า threshold เป็นตัวกำหนดว่าน้ำหนักรวมของ ข้อมูลเข้าต้องมากขนาดไหนจึงจะสามารถส่งข้อมูลออกไปยัง นิวรอนตัวอื่นได้ เมื่อนำ นิวรอนแต่ละ หน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกจะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลข เพราะฉะนั้นเมื่อนำการทำงานเข้ามาใช้กับคอมพิวเตอร์ สามารถเขียนได้ดังนี้

$$\text{if (sum(input * weight) > threshold) then output}$$

เมื่อมี input แทนข้อมูลเข้า เข้ามายังโครงข่าย จะใช้ข้อมูลเข้า คูณกับ weight แทนน้ำหนักของ นิวรอนแต่ละขา ผลที่ได้จากข้อมูลเข้าทุกขาของนิวรอน จะรวมกันแล้วก็เอามาเทียบกับ threshold ที่ กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้วนิวรอนจะส่งข้อมูลออกไปข้อมูลออกนี้ก็จะถูก ส่งไปยังข้อมูลเข้า ของนิวรอนอื่น ๆ ที่เชื่อมกันในโครงข่ายถ้าค่าน้อยกว่า threshold ก็จะไม่เกิดข้อมูล ออกดังแสดงในภาพที่ 2.12

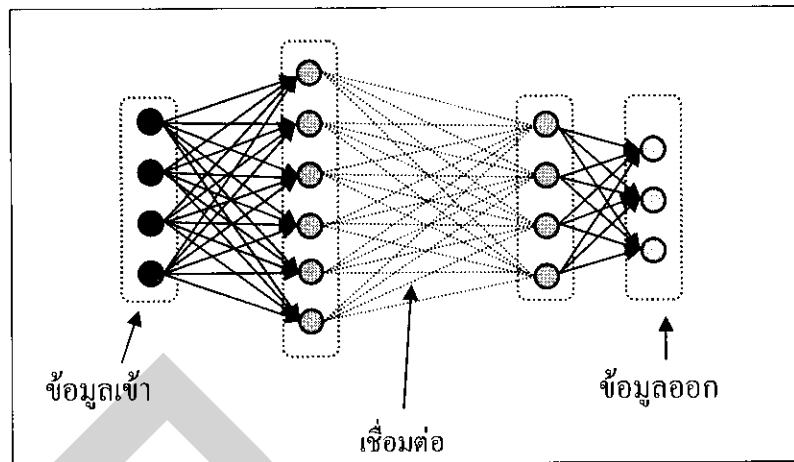


ภาพที่ 2.12 ตัวแบบของนิวรอนในคอมพิวเตอร์

สถาปัตยกรรมโครงข่ายประสาทเทียมที่พบทั่วไปจะมีลักษณะหลัก ๆ คือ มีการจัดเซลล์ประสาทเทียมเป็นชั้น (layer) ชั้นที่รับข้อมูลเข้าเรียกว่า ชั้นข้อมูลเข้า (input layer) ชั้นที่ผลิตผลตอบของโครงข่ายเรียกว่า ชั้นข้อมูลออก (output layer) ส่วนชั้นอื่น ๆ ที่มีส่วนในการช่วยทำการประมวลผลอยู่ภายในเรียกว่า ชั้นซ่อน (hidden layer) ในโครงข่ายประสาทเทียมอาจมีชั้นซ่อนได้หลายชั้นโครงสร้างพื้นฐานจะมีลักษณะเป็นการประกอบกันของรูปแบบ ดังต่อไปนี้

1. แบบป้อนไปข้างหน้า

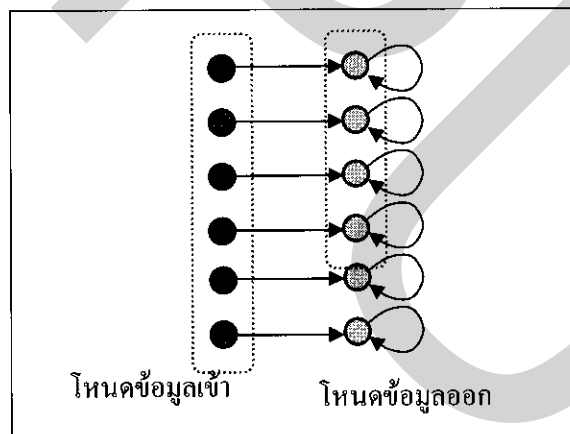
แบบป้อนไปข้างหน้า (feedforward network) ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจาก โหนดข้อมูลเข้า ส่งต่อมาเรื่อย ๆ จนถึง โหนดข้อมูลออก โดยไม่มีการย้อนกลับของข้อมูล หรือแม้แต่ โหนดในชั้นเดียวกันก็ไม่มี การเชื่อมต่อกัน สามารถจัดได้เป็น 2 แบบคือ แบบมีชั้นของเซลล์ประสาทชั้นเดียว และแบบมีชั้นของเซลล์ประสาทหลายชั้น โดยปกติการเชื่อมโยงจะถูกกำหนดขึ้นระหว่างชั้นที่ติดกัน โดยจะมีการเชื่อมโยงระหว่างเซลล์ประสาทเทียมทุกตัว จากชั้นหนึ่ง ๆ ไปยังเซลล์ประสาทเทียมทุกตัวในชั้นต่อไป ในบางสถาปัตยกรรมอาจมีการเชื่อมโยงข้ามชั้นก็ได้ ดังแสดงในภาพที่ 2.13



ภาพที่ 2.13 สถาปัตยกรรมแบบป้อนไปข้างหน้า

2. แบบมีการป้อนไปเวียนกลับ

แบบมีการป้อนไปเวียนกลับ (feedback network) บางที่เรียกว่า recurrent network ข้อมูลที่ประมวลผลโครงข่ายประสาทเทียมจะมีการป้อนกลับเข้าไปยังวงจรซ้ำหลายๆ ครั้ง จนกระทั่งได้คำตอบออกมา จะมีการเชื่อมโยงที่ถูกกำหนดขึ้นระหว่างเซลล์ประสาทเทียมในชั้นหนึ่งๆ ย้อนกลับไปยังชั้นอื่นๆ ก่อนหน้านั้น หรือแม้แต่ภายในชั้นเดียวกันเอง ดังแสดงในภาพที่ 2.14

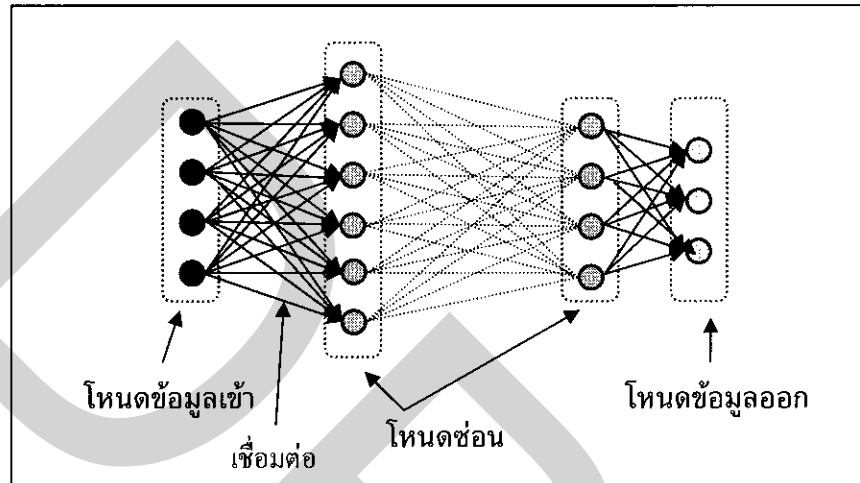


ภาพ 2.14 สถาปัตยกรรมแบบมีการป้อนไปเวียนกลับ

3. แบบป้อนไปข้างหน้าหลายชั้น

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น (multiple feedforward) ใช้การเชื่อมโยงแบบป้อนไปข้างหน้าประกอบด้วย 3 ชั้น ได้แก่ ชั้นโหนดข้อมูลเข้า ที่ถูกเชื่อมต่อกับชั้นของโหนดซ่อน และจะมีการเชื่อมต่อกับชั้นของโหนดข้อมูลออก เป็นการเพิ่มจำนวนชั้นใน ชั้นซ่อนดังแสดง

ในภาพที่ 2.13 การทำงานของโหนดข้อมูลเข้าจะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่ถูกป้อนเข้าสู่เครือข่าย และการทำงานของแต่ละ โหนดซ่อน จะถูกกำหนด โดยการทำงานของโหนดข้อมูลเข้าและค่าน้ำหนักบนความสัมพันธ์ระหว่าง โหนดข้อมูลเข้า และ โหนดซ่อน พฤติกรรมการทำงานของ โหนดข้อมูลออก จะขึ้นอยู่กับการทำงานของ โหนดซ่อน และค่าน้ำหนักระหว่าง โหนดซ่อน และโหนดข้อมูลออก ดังแสดงในภาพที่ 2.15



ภาพ 2.15 โครงสร้างแบบป้อนไปข้างหน้าหลายชั้น

ลักษณะการทำงานแบบนี้สามารถกำหนดการแทนค่าให้แก่ โหนดข้อมูลเข้าได้อย่างอิสระ ค่า น้ำหนักระหว่าง โหนดข้อมูลเข้าและ โหนดซ่อน จะถูกกำหนดเมื่อ โหนดซ่อน กำลังทำงาน ฉะนั้นเวลาที่แก้ไขค่าน้ำหนัก โหนดซ่อน จะสามารถเลือกได้ว่าอะไรคือค่าที่เราแทนเข้ามาในโมเดลทางคณิตศาสตร์ของโครงข่ายประสาทเทียม แบบป้อนไปข้างหน้าหลายชั้น แสดงในภาพ มีเวกเตอร์อินพุต $X = \{x_i\}$ กำหนดให้ $x_i \in \mathbb{R}^p$ และเวกเตอร์ผลลัพธ์ $Y = \{y_i\}$ กำหนดให้ $y_i \in \mathbb{R}^m$ ซึ่งประกอบด้วยสมการการคำนวณภายในเซลล์ประสาทแต่ละชั้นดังนี้

ชั้นซ่อน

$$h_j = \sigma(u_j)$$

$$u_j = \sum_{i=1}^p w_{ji}^h x_i$$

กำหนดให้ h_j คือผลลัพธ์ของเซลล์ใน ชั้นซ่อน ซึ่ง $1 \leq j \leq L$

$\sigma(\cdot)$ คือ activation function

w_{ji}^h คือ ค่าของน้ำหนักบนเส้น ji เชื่อมเซลล์ในชั้นซ่อน j กับเซลล์ในชั้นข้อมูลเข้า i

W^h คือ เวกเตอร์น้ำหนัก $\{w_{ji}^h\}$

ชั้นข้อมูลออก

$$y_k = \sigma(v_k),$$

$$y_k = \sum_{j=1}^m w_{kj}^y h_j$$

โดยที่ y_k คือ ผลลัพธ์ของเซลล์ใน ชั้นข้อมูลออก ซึ่ง $1 \leq k \leq m$

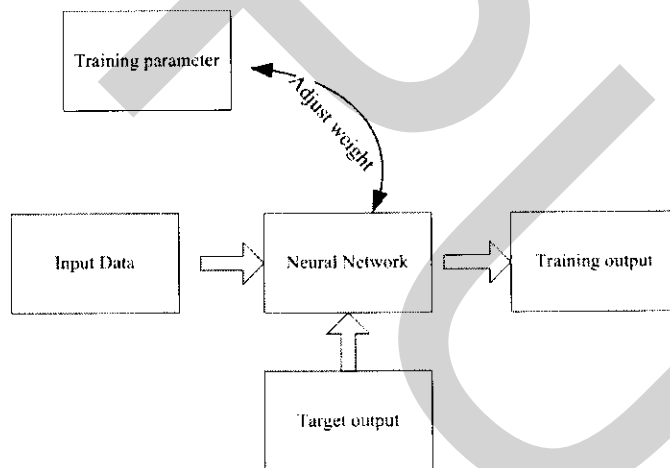
$\sigma(\cdot)$ คือ activation function

w_{ki}^h คือ ค่าของน้ำหนักบนเส้น ki เชื่อมเซลล์ในชั้นซ่อน k กับเซลล์ในชั้นข้อมูลเข้า j

W^y คือ เวกเตอร์น้ำหนัก $\{w_{kj}^y\}$

▪ รูปแบบการเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้เป็นกระบวนการที่สามารถทำให้เปลี่ยนแปลงพฤติกรรม ความคิด สามารถเรียนรู้ได้จากการ
ได้ยีนการสัมผัส การอ่าน การใช้เทคโนโลยี สามารถแบ่งออกได้ 2 แบบ



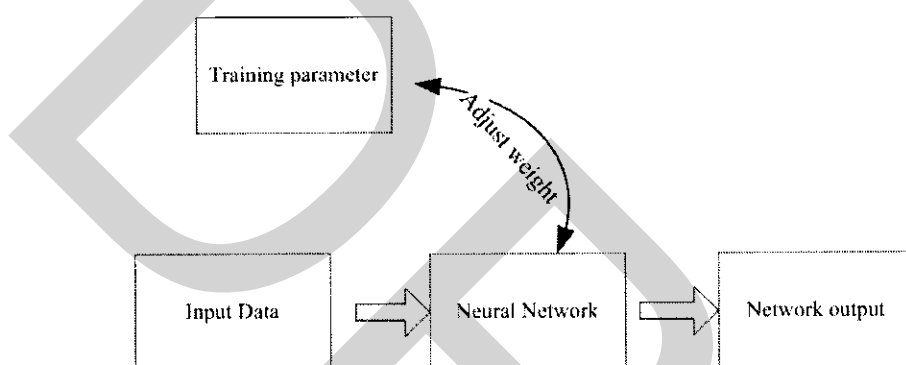
ภาพที่ 2.16 การเรียนรู้แบบมีการสอน

1. การเรียนรู้แบบมีการสอน

การเรียนรู้แบบมีการสอน (supervised learning) เป็นการเรียนรู้แบบที่มีการตรวจคำตอบ
เพื่อให้งจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจดูว่าวงจรข่ายให้คำตอบที่
ถูกหรือไม่ ถ้าตอบไม่ถูก การเรียนรู้โดยให้ข่ายงานปรับตัวเข้ากับค่าจริง (target) หรือคำตอบที่มีอยู่
เพื่อให้ได้คำตอบที่เหมือนหรือใกล้เคียงกับคำตอบจริงดังแสดงในภาพที่ 2.16

2. การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด จะไม่มีค่าจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้างของตัวเอง ตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่น การแยก ลักษณะของคน พิซ และสัตว์ เองได้โดยไม่มีผู้สอนจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้าง ของตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่นการ ที่เราสามารถแยกลักษณะของคน พิซ และสัตว์ เองได้โดยไม่มีผู้สอนดังแสดงในภาพที่ 2.17



ภาพที่ 2.17 การเรียนรู้แบบไม่มีผู้สอน

▪ การเรียนรู้แบบแพร่ย้อนกลับ

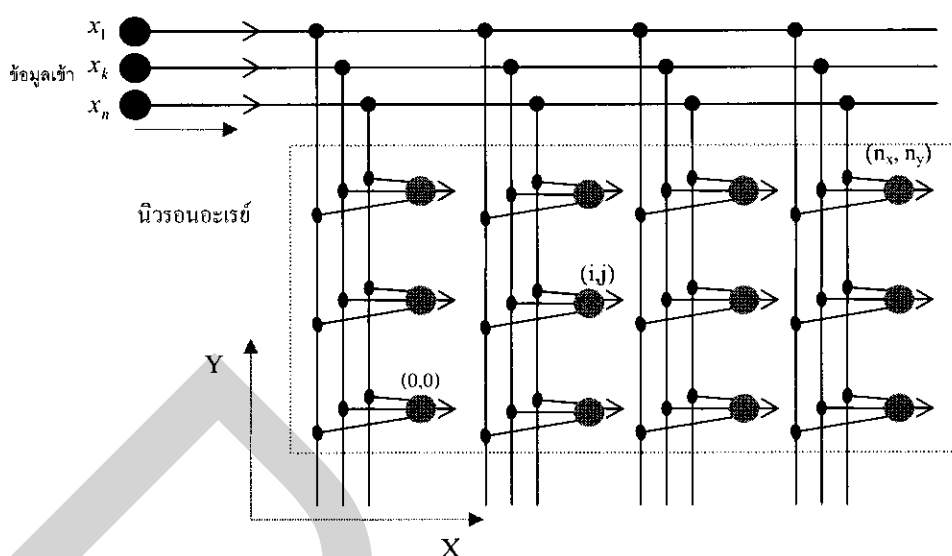
การเรียนรู้แบบแพร่ย้อนกลับ (back-propagation learning) ถูกนำไปประยุกต์ใช้มากที่สุด ในการใช้งานในโครงข่ายประสาทเทียม โดยลักษณะของการเรียนรู้แล้ว การเรียนรู้แบบแพร่ย้อนกลับ ไม่จำเป็นต้องใช้เฉพาะโครงข่ายประสาทเทียมที่มีสถาปัตยกรรมแบบป้อนไปข้างหน้าหลายชั้นเพียง แบบเดียว แต่การจับคู่ดังกล่าวได้รับความนิยมนำไปใช้งานมากที่สุด การเรียนรู้แบบแพร่ย้อนกลับนี้ เป็นการเรียนรู้แบบมีการสอน (supervise learning) ขั้นตอนในการเรียนรู้แบบแพร่ย้อนกลับจะเริ่ม ด้วยการขนส่งข้อมูลที่จะใช้สอนเข้าไปในเครือข่ายทางชั้นข้อมูลเข้า ชั้นข้อมูลเข้าซึ่งโดยปกติจะไม่มี การประมวลผล จะทำหน้าที่ส่งกระจายข้อมูลนั้นไปยังเซลล์ประสาทเทียมต่างๆในชั้นซ่อน จากนั้นเซลล์ ประสาทเทียมทุกตัวในชั้นซ่อน จะทำการประมวลผลตามวิธีการของแบบจำลองเซลล์ประสาทเทียม การประมวลผลเกิดจากชั้นซ่อนชั้นแรกผ่านไปตามชั้นซ่อนต่างๆจนสร้างผลตอบที่ชั้นข้อมูลออก

ผลที่ตอบออกมาจากชั้นเอาต์พุตจะถูกเปรียบเทียบกับค่าเป้าหมายที่กำหนด เพื่อคำนวณค่าความคลาดเคลื่อน (error) จากนั้นปริมาณการปรับแต่งค่าน้ำหนักของแต่ละชั้นสัญญาณเข้าของแต่ละเซลล์ประสาทเทียมในชั้นข้อมูลออกก็จะถูกคำนวณขึ้น โดยดูจากความคลาดเคลื่อนและปริมาณการปรับค่าน้ำหนักที่ชั้นสัญญาณของแต่ละเซลล์ประสาทเทียมในชั้นซ่อนต่างๆ แต่เนื่องจากไม่สามารถกำหนดค่าเป้าหมายของผลตอบจากเซลล์ประสาทเทียมในชั้นซ่อนจึงถูกมองว่า เป็นการไหลย้อนกลับของความคลาดเคลื่อนของเซลล์ประสาทเทียมต่างๆ ของชั้นข้อมูลออกหลังจากรวบรวมค่าความคลาดเคลื่อนที่ไหลย้อนกลับมาแล้ว หลักการเดียวกับชั้นข้อมูลออก คือ ดูจากผลรวมความคลาดเคลื่อนที่ไหลย้อนกลับมาไปยังเซลล์ประสาทเทียมตัวนั้น ๆ ประกอบกับค่าของสัญญาณเข้าที่วิ่งผ่านเข้ามาทางชั้นสัญญาณเข้านั้นด้วย

เมื่อค่าน้ำหนักที่ชั้นสัญญาณเข้าของเซลล์ประสาทเทียมทุกตัวถูกปรับแต่งหมดแล้วโครงข่ายประสาทเทียมก็พร้อมจะเรียนรู้ข้อมูลตัวอย่างต่อไป ในทางปฏิบัติแล้ว จะต้องใช้ข้อมูลตัวอย่างจำนวนมาก และจะต้องสอนข้อมูลตัวอย่างเหล่านั้นซ้ำกันหลายรอบ จึงจะสามารถสอนโครงข่ายประสาทเทียมชนิดนี้จนเรียนรู้ได้อย่างถูกต้อง

- ข้อดีโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) ซึ่งเลียนแบบการทำงานและคุณสมบัติเซลล์สมองหรือระบบประสาทของมนุษย์ เมื่อโครงข่ายประสาทเทียมผนวกกับความสามารถของวิทยาการคอมพิวเตอร์ในปัจจุบัน เช่น หน่วยความจำ การประมวลผลที่รวดเร็ว แม่นยำ และค่าใช้จ่ายที่ไม่สูงนัก ทำให้ได้ระบบที่มีศักยภาพในการทำงาน มีคุณลักษณะและคุณสมบัติที่น่าสนใจ เช่น สามารถจำลองปัญหาได้โดยไม่ต้องทราบรูปแบบการกระจายของข้อมูล (distribution free) มีข้อผิดพลาดได้บ้าง (fault tolerance) เรียนรู้ด้วยตนเองได้ (self-organization) มีการทำงานแบบขนาน (parallel process) รวดเร็วระบบทำงานโดยใช้เพียงฟังก์ชันทางคณิตศาสตร์อย่างง่ายแทนที่จะใช้กลไกทางชีวเคมี และไม่ได้ทำงานตามชุดคำสั่งแต่อย่าง เดียวดังเช่นโปรแกรมคอมพิวเตอร์ทั่วไป คำตอบหรือผลลัพธ์น่าเชื่อถือ ด้วยเหตุผลดังกล่าวโครงข่ายประสาทเทียมจึงสามารถแก้ปัญหาใกล้เคียงกับเซลล์สมองหรือระบบประสาทของสิ่งมีชีวิตโดยเฉพาะมนุษย์ ระบบเรียนรู้หรือรู้จำจากตัวอย่างที่มีจำนวนและความหลากหลาย แหล่งที่มาของตัวอย่างอาจได้จากการสัมภาษณ์ตรง (direct interview) ข้อมูลในอดีต (historical record) หรือกระบวนการจำลอง (simulation)



ภาพที่ 2.18 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง

2.3.2 ทฤษฎีแผนผังการจัดระเบียบตัวเอง

ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [T. Kohonen, 1995] [T. Mitchell, 1997] เป็นอัลกอริทึมนิเวรอนเน็ตเวิร์คที่นิยมใช้มากที่สุด โดยแนวคิดของ SOM คือ การทำซ้ำข้อมูลเพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ เป็นวิธีเกี่ยวกับการจัดกลุ่มด้วยตัวเองโดยใช้โครงสร้างตาข่ายระบบประสาท เป็นการเรียนรู้แบบไม่มีผู้สอน มีลักษณะการเรียนรู้ แบบเครือข่าย 2 ชั้น เป็นการจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติคือการจัดกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้เคียงกัน โดยโครงสร้างการทำงานของ SOM ชั้นแรกทำหน้าที่นำเข้าข้อมูลและจัดส่งข้อมูลให้แก่นิเวรอนชั้นที่สอง ทุกโหนดในตาราง หรือ อะเรย์นิเวรอน (array of neurons) ระหว่างชั้นจะเชื่อมต่อกันด้วยค่าน้ำหนัก (weight) จากนั้นข้อมูลจะถูกส่งไปยังนิเวรอนในชั้นที่สอง เพื่อทำการเปรียบเทียบว่าใกล้เคียงกับค่ากลางกลุ่มใดมากที่สุด แต่ละโหนดในชั้นนี้จะมีความสัมพันธ์กันแบบ เพื่อนบ้าน (neighborhood relation) ทำให้เกิดเป็นรูปแบบ 2 มิติ จากนั้นจะทำการปรับค่า น้ำหนักของตัวที่เป็นผู้ชนะ (winner) ทำให้ข้อมูลที่ข้อมูลเข้า เข้ามาเกิดการปรับเปลี่ยน และเมื่อผ่านการเรียนรู้ไปหลายๆรอบจะทำให้ ได้กลุ่มข้อมูลออกมาเป็นผลลัพธ์ จากนั้นจึงนำผลลัพธ์ที่ได้นั้นมาผ่านกระบวนการ Visualization เพื่อแสดงผลลัพธ์ที่ได้นั้นออกมาเป็นกราฟชนิดต่าง ๆ ต่อไป

ขั้นตอนการทำงานของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

1. ทำการจัดกลุ่มของข้อมูลนำเข้าให้อยู่ในกลุ่มข้อมูลที่ใกล้เคียง ทำการ Normalize ค่าข้อมูลแต่ละตัว เพื่อให้ได้ค่ากลางเป็น $x_i = \frac{x_i - \min}{\max - \min}$, กำหนดให้ x_1, x_2, \dots, x_n เป็นข้อมูลที่ถูกนำเข้า n คือจำนวนตัวอย่างข้อมูลเข้า \min คือ ค่าที่น้อยที่สุดของชุดข้อมูล \max คือ ค่าที่มากที่สุดของชุดข้อมูล
2. กำหนดโครงสร้างโครงข่ายประสาทเทียมในลักษณะสองมิติ กำหนด แนวแกน x และแกน y กำหนดชุดของข้อมูล สุ่มค่าน้ำหนักเริ่มต้น ในโครงข่ายดังแสดงในภาพที่ 2.18
3. สุ่มค่าเริ่มต้นให้กับค่ากลางของกลุ่ม (cluster center) $W = \{w_1, w_2, \dots, w_p\}$, โดยที่ p คือ จำนวนกลุ่ม เวกเตอร์ค่ากลางคือ $W_j = \{w_{j1}, w_{j2}, \dots, w_{jm}\}$, เมื่อ $1 \leq j \leq p$
4. กำหนดหาระยะทางระหว่างโครงข่ายโดย ยูคลิดีเนียน (Euclidean distance) เพื่อหาผู้ชนะ (winner) ซึ่งจะหาได้จากโครงข่ายที่ใช้ระยะทางที่ได้จากการคำนวณ เพื่อหาค่าน้อยที่สุด

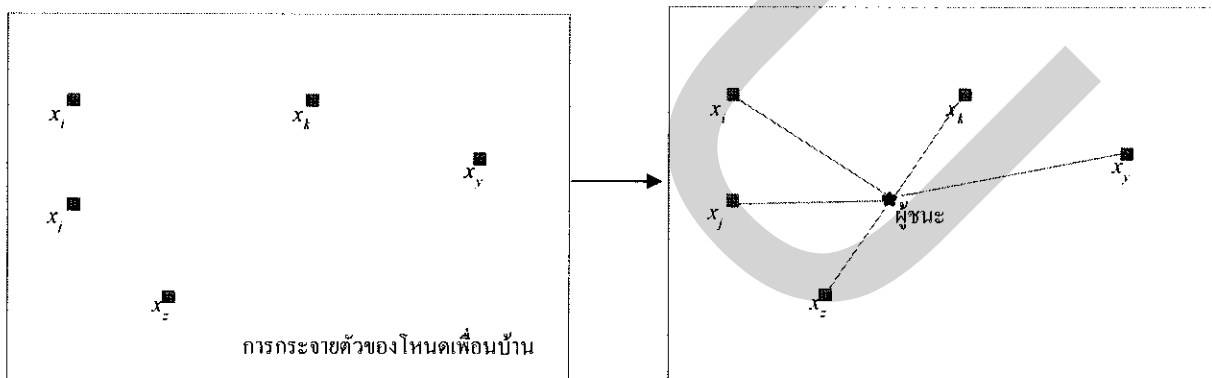
$$C(k_1, k_2) = \min_{i,j} C_{i,j}$$

เมื่อ k_1, k_2 คือดัชนีของโหนดผู้ชนะ

$$C_{i,j}^2 = \|x - w_j\|^2$$

$$C_{i,j}^2 = \sum_{i=1}^m (x_i - w_{ij})^2$$

เมื่อ $C_{i,j}$ ค่าความต่างระหว่างข้อมูลนำเข้า x_i กับเวกเตอร์น้ำหนัก $w_{i,j}$



ภาพที่ 2.19 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน

5. คำนวณค่ากลางสำหรับกลุ่มที่เป็น ค่าผู้ชนะใหม่

$$w_j(t+1) = w(t) + \eta(x_i - w_j(t))$$

6. ทำซ้ำขั้นตอนที่ 4-5 จนครบชุดข้อมูลนำเข้า หรือจนกว่าค่าน้ำหนักเริ่มจะคงที่
7. โหนดที่เป็นผู้ชนะจะสอดคล้องตามสูตร $C(k_1, k_2) = \min_{i,j} C_{i,j}$ เมื่อ k_1, k_2 ดัชนีของโหนดที่เป็นผู้ชนะ

8. โหนดที่เป็นโหนดเพื่อนบ้านจะถูกกำหนดโดย $h(\rho, t) = \exp\left(-\frac{\rho^2}{2\sigma^2(t)}\right)$

9. ทหาระยะห่างระหว่างโหนดนั้น ๆ กับ โหนดที่เป็นผู้ชนะ $\rho = \sqrt{(k_1 - i)^2 + (k_2 - j)^2}$

10. หาโหนดเพื่อนบ้านที่ใกล้เคียงที่สุด โดยขอบเขตจะลดลงตามเวลาตามรูปแบบดังนี้

$$h(\rho, t) = \exp\left(-\frac{\rho^2}{\sigma^2(t)}\right) \left(1 - \frac{2}{\sigma^2(t)} \rho^2\right)$$

$$h(\rho) = \begin{cases} 1, & |\rho| \leq a, \\ \frac{1}{3}, & a < |\rho| \leq 3a, \\ 0, & |\rho| > 3a, \end{cases}$$

11. ปรับค่าน้ำหนัก ของแต่ละโหนดด้วย

$$W_{i,j}(t+1) = W_{i,j}(t) + a(t)h(\rho, t)(X^i(t) - W(t))$$

- ข้อเสียของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

การเรียนรู้นี้ไม่ทราบค่าที่ต้องการจึงไม่สามารถจะหยุดการเรียนรู้เมื่อไร ต้องมีการปรับค่าน้ำหนักไปเรื่อย ๆ ปัญหานี้สามารถแก้ไขได้ โดยอาศัยเงื่อนไขหรือหลักเกณฑ์บางอย่างในการใช้เป็นเครื่องมือในการหยุดการเรียนรู้

ค่ากลางบางตัวซึ่งเกิดจากการสุ่ม อาจจะถูกในบริเวณที่ห่างจากกลุ่มค่ากลางตัวอื่นมาก ซึ่งไม่สามารถเป็น ตัวชนะได้เลย จึงต้องเพิ่มกระบวนการในการแก้ปัญหา

ไม่สามารถ กำหนดจำนวนชุดข้อมูลที่นำมาจัดกลุ่ม ควรมีที่กลุ่มถึงจะเหมาะสม จึงควรมีการศึกษาวิจัยต่อไป

2.3.3 ทฤษฎีความเชื่อเบย์

ทฤษฎีความเชื่อเบย์ (ทฤษฎีความเชื่อเบย์ (bayesian approach) [R. O. Duda, 1973] [T. Mitchell, 1997] เป็นการใช้รูปแบบของการประมาณค่าของความน่าจะเป็นที่จะเกิดเหตุการณ์ โดยอาศัยความสัมพันธ์ที่เกิดขึ้นของข้อมูลเพื่อมาคำนวณเหตุการณ์ที่เกิดขึ้น ดังนั้นทฤษฎีความเชื่อเบย์จึงอยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่จะเกิดขึ้น ในเบื้องต้นจะกล่าวถึงทฤษฎีพื้นฐานของความน่าจะเป็น (basic formulas for probability) ดังนี้ ถ้ากำหนดให้ A และ B เป็นเหตุการณ์ที่เป็นอิสระต่อกัน คือเหตุการณ์ A และ B ไม่มีความเกี่ยวข้องกัน สามารถเขียนสมการได้

$$p(A \cap B) = p(A) \cdot p(B)$$

โดยที่ $p(A \cap B)$ หมายถึงความน่าจะเป็นที่เหตุการณ์ A และ B เกิดขึ้นพร้อมกัน แต่ถ้ากำหนดให้เหตุการณ์ A และ B เป็นเหตุการณ์ที่ขึ้นต่อกัน

$$p(A \cap B) = p(A) \cdot p(B|A)$$

$$p(A \cap B) = p(B) \cdot p(A|B)$$

$$p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

เมื่อแทนค่าความน่าจะเป็นของ A โดยที่เหตุการณ์ B ได้เกิดขึ้นแล้ว จึงสามารถเขียนสูตรของเบย์ ได้ดังนี้

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

- วิธีการเรียนรู้เบย์อย่างง่าย

วิธีการเรียนรู้เบย์อย่างง่าย (naïve-bayes learning) [T. Mitchell, 1997] วิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายก็เพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วยความรู้ก่อนหน้า หมายถึง ความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่เราจะเก็บข้อมูล เมื่อใช้งานเรานำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง เมื่อ C แทนกลุ่มข้อมูลที่จะถูกแบ่งประกอบด้วยกลุ่มที่ถูกสังเกต $c \in \{X_1, X_2, \dots, X_k\}$ แทนแอตทริบิวต์เวกเตอร์ข้อมูล ประกอบด้วยข้อมูลที่สังเกต $x = (x_1, x_2, \dots, x_k)$ สามารถเขียน $X = x$ โดยย่อได้ดังนี้

$$X = x_1 \wedge x_2 \wedge \dots \wedge x_k$$

สำหรับค่าผิดพลาดของการทำนายเป็น

$$\operatorname{argmax}_c (p(C = c|X = x))$$

สามารถเขียนสมการเบย์ได้ดังนี้

$$p(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(X = x)},$$

$$p(C = c|X = x) \propto p(C = c)p(X = x|C = c),$$

โดยที่ค่าของ $p(C = c)$ และ $p(X = x|C = c)$ จะถูกประมาณค่าจากการเรียนรู้ซึ่งในความเป็นจริงแล้วไม่สามารถที่จะประมาณค่าของ $p(X = x|C = c)$ ได้โดยตรง ดังนั้นเมื่อค่าของ X_1, X_2, \dots, X_k เป็นเหตุการณ์ที่สามารถเกิดขึ้นต่อกัน แล้ว

$$p(X = x|C = c) = p(\bigwedge_{i=1}^k X_i = x_i|C = c)$$

$$p(X = x|C = c) = \prod_{i=1}^k p(X_i = x_i|C = c)$$

โดยทั่วไปนิยมเขียนเป็น

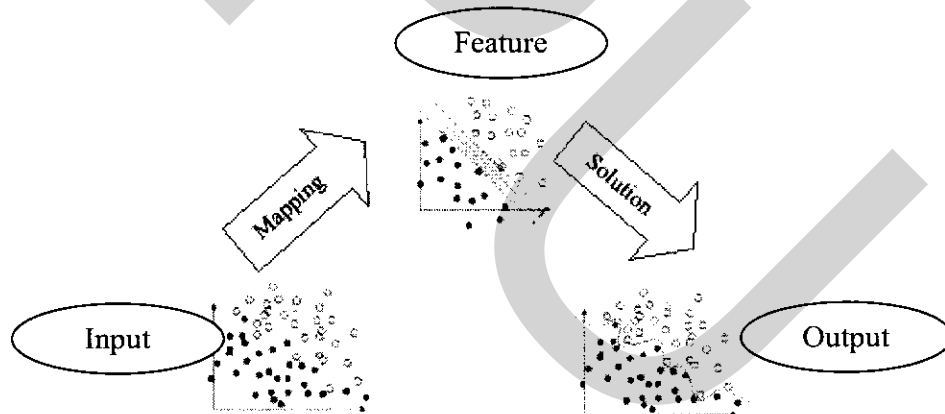
$$p(C = c|X = x) \propto p(C = c) \prod_{i=1}^k p(X_i = x_i|C = c)$$

- ข้อดีของวิธีการเรียนรู้แบบเบย์

เบย์ เป็นวิธีการเรียนรู้ โดยใช้ ความรู้ก่อนหน้า (prior knowledge) เข้ามาช่วยในการเรียนรู้ได้ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น สามารถลดข้อจำกัดอย่างง่ายในสมมุติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัติ แต่ในความเป็นจริงคุณสมบัติบางตัวจะขึ้นต่อกัน และควรนำค่าความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดล จึงใช้ข่างานเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปร เพื่อทำให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่างานความเชื่อเบย์ให้อยู่ในรูปโครงสร้างข่างานและตารางความน่าจะเป็นแบบมีเงื่อนไข

2.3.4 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) [R. O. Duda 1973] เป็นตัวแบบที่ใช้ในการแยกแยะข้อมูล โดย SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่มชุดข้อมูล โดยตัวแบบของ SVM มีเกี่ยวข้องกับโครงข่ายประสาทเทียม ตัวแบบของ SVM ใช้ sigmoid kernel function ซึ่งมีค่าเท่ากันทั้ง 2 ชั้น (layer) ตัวแบบของ SVM มีความคล้ายคลึงกับเพอร์เซพตรอน (perceptron) ซึ่งเป็นข่ายงานประสาทเทียมแบบง่ายมีหน่วยเดียวที่จำลองลักษณะของเซลล์ประสาทด้วยการใช้ kernel function ซึ่ง SVM จะใช้ข้อมูลของ คุณสมบัติและตัวแปรที่เปลี่ยนแปลงใช้ในการกำหนดระนาบหลายมิติ เรียกว่า โครงสร้าง (feature) ส่วนการเลือกที่มีความเหมาะสมที่สุดเรียกว่า โครงสร้างในการคัดเลือก (feature selection) จำนวนเซตของโครงสร้างที่ใช้อธิบายในกรณีหนึ่ง (เช่น แถวของการค่าคาดการณ์) เรียกว่า เวกเตอร์(vector) ดังนั้นจุดมุ่งหมายของตัวแบบ SVM คือการประโยชน์สูงสุดจากระนาบหลายมิติที่แบ่งแยกกลุ่มของเวกเตอร์ที่ถูกเรียกว่า การหาเส้นแบ่ง hyper-planes ซึ่งใช้แบ่งข้อมูลสองคลาสเพื่อให้ได้ผลลัพธ์ที่ดี โดยพิจารณาจากสมการเส้นตรง hyper planes และ SVMs จะทำการค้นหาเวกเตอร์ที่อยู่ใกล้เส้นแบ่ง hyper planes เรียกว่า ซัพพอร์ตเวกเตอร์(support vectors) ดังแสดงในภาพที่ 2.20



The SVM algorithm

ภาพ 2.20 หลักการทำงานของ SVM

▪ หลักการทำงานของ SVM

1. ข้อมูลที่อินพุตเข้าคำนวณหาค่า y ซึ่งค่าของ $y \in \{-1,1\}$ ได้จากสมการ $y = w^T \cdot x + b$ ถ้าค่าของ $w^T \cdot x + b > 0$ จะกำหนดให้ค่า $y = 1$ ซึ่งจะจัดอยู่ใน class ที่ 1 ถ้าค่าของ $w^T \cdot x + b < 0$ จะกำหนดให้ค่า $y = -1$ ซึ่งจะจัดอยู่ใน class ที่ 2

2. คำนวณหาเส้นตรงที่แบ่งเอกสารซึ่งเรียกว่า เส้น optimal hyper plane จากสมการ

$$w^T \cdot x + b = 0$$

3. นำค่าที่ได้จากข้อที่ 1 และ 2 ไปเขียนบนเส้นตรงตามแนวแกนตั้งและแกนนอนจะได้ดังภาพที่ 9 โดยระยะทาง (d) หรือ maximum margin จากเส้นขอบ ณ จุด x_i ไปยัง hyper plane สามารถแสดงได้ดังสมการ

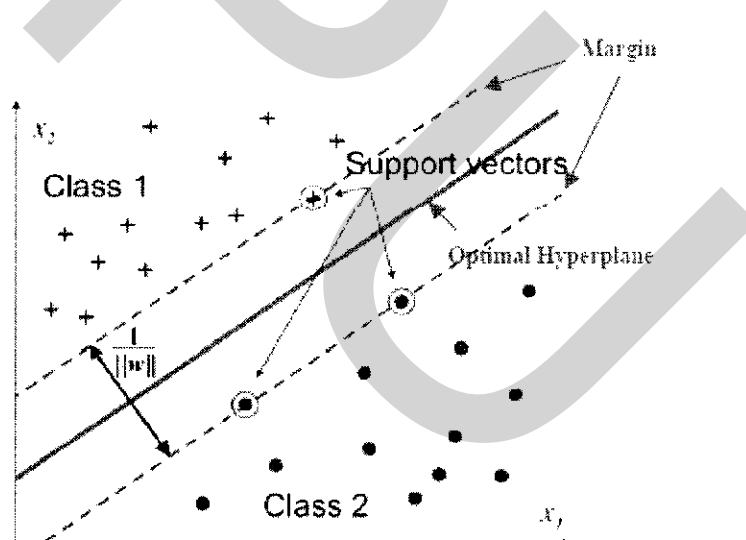
$$d = \frac{|w^T \cdot x + b|}{\|w\|}$$

โดยกำหนดให้

w คือ เวกเตอร์น้ำหนัก (weight vector)

x_i คือ input vector ของเอกสาร

b คือ ค่าคงที่ที่กำหนดขึ้นเพื่อให้เหมาะสมกับการจัดกลุ่มเอกสาร



ภาพ 2.21 การแบ่งข้อมูลโดย SVMs

4. เลือกจุดที่อยู่ใกล้เส้นตรง optimal hyper plane ทั้งเหนือเส้นซึ่ง เรียกว่า “ขอบล่าง” ซึ่งเป็นขอบล่างสุดของ class เอกสารที่อยู่เหนือเส้นตรง optimal hyper plane และใต้เส้นเรียกว่า “ขอบบน” ซึ่งเป็นขอบบนสุดของ class เอกสารที่อยู่ใต้เส้นตรง optimal hyper plane เพื่อที่จะหา ระยะทางระหว่างเส้นขอบทั้งสองโดยจะเลือกเอาค่าระยะทางที่ห่างจากเส้นตรง optimal hyper plane ที่น้อยที่สุดเป็นตัวเลือกในการจัดกลุ่มเอกสาร

อย่างไรก็ตามโดยพื้นฐานของ SVMs นั้น จะสามารถแบ่งกลุ่มข้อมูลได้เพียง 2 กลุ่ม ดังแสดง ในภาพที่ 2.21 ดังนั้นการปรับเทคนิคของการเรียนรู้ด้วย SVMs เพื่อให้ได้เป็นการจัดแบบหลายกลุ่ม จึงเป็นสิ่งจำเป็น สำหรับงานวิจัยฉบับนี้จะปรับปรุงขั้นตอนการเรียนรู้ โดยเป็นการสร้างโมเดลการจัด กลุ่มด้วย One Class SVMs นั่นคือ ให้แต่ละกลุ่มข้อมูลที่กำลังสนใจนั้นเป็น $w^T \cdot x + b > 0$ โดยค่า $y = 1$ เมื่อผ่านข้อมูลแต่ละชุด (ที่ผ่านการจัดกลุ่มด้วยมือไว้ก่อนหน้า) เข้าสู่กระบวนการเรียนรู้ ก็จะ สร้างโมเดลของการจัดกลุ่มเอกสารแต่ละกลุ่มด้วย SVMs นอกจากนี้การจัดกลุ่มเอกสารด้วย SVMs หากต้องการความถูกต้องอย่างมาก จะต้องคำนึงถึงลักษณะของข้อมูลที่มี เพื่อให้สามารถเลือก kernel function ของการทำงานได้อย่างเหมาะสม เพราะ kernel function จะเป็นปัจจัยในการทำงานที่สำคัญของ SMV โดยทั่วไป kernel function ที่ใช้งานกับ SVMs มี 4 ประเภทคือ Linear , Polynomial , Radial Basis Function และ Sigmoid

ด้วยแนวคิดของการจัดกลุ่มด้วย SVM จะเป็นการสร้าง hyperplane เพื่อแยกกลุ่มเป็น 2 กลุ่ม แบบที่ต้องดูค่า maximum margin ที่เหมาะสมที่สุดในการจัดกลุ่ม การใช้ maximum margin ตามทฤษฎีของ Vapnik Chervonenkis ด้วยการดูค่าความผิดพลาดที่น้อยที่สุดเมื่อได้ค่า Margin ที่มากที่สุด แต่การทำลักษณะเช่นนี้แม้จะมีประโยชน์ แต่อาจจะก่อให้เกิดเส้นแบ่งเขตแดนที่ไม่เหมาะสม เพราะเกิดค่าความ ผิดพลาดสูง ดังนั้นจึงต้องกำหนดพารามิเตอร์ของค่า maximum margin hyper plane ดังนั้นจึงต้องมีการนำการทำงานอื่น ๆ เข้ามาช่วย ซึ่งก็คือการนำเอาหลักการเรื่อง kernel เข้ามาใช้เพื่อช่วยในการหาค่า maximum margin hyper plane

2.4 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการ ทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมี

การทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และ F-measure โดยยกตัวอย่างของค่าที่เกิดขึ้นจากตารางที่ 2.1

		ค่าทำนาย (predicted)	
		ปฏิเสธ (false/negative)	ยอมรับ (true/positive)
ค่าความจริง (actual)	ปฏิเสธ (negative)	a	b
	ยอมรับ (positive)	c	d
ค่าความถูกต้อง (accuracy)		Acc	

ตารางที่ 2.1 การวัดประสิทธิภาพ

- ค่าความแม่นยำ (false positive rate / Precision: Pr) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหาได้

$$Pr = \frac{b}{a+b}, a+b > 0$$

- ค่าความระลึก (true positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพที่ถูกต้องทั้งหมด

$$Re = \frac{b}{b+c}, b+c > 0$$

- ค่าความถูกต้อง (accuracy: Acc) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องทั้งหมดจากจำนวนภาพที่มีอยู่

$$Acc = \frac{(a + d)}{(a + b + c + d)}$$

- ค่า F-measure เป็นการวัดค่าความสัมพันธ์ระหว่างค่าความระลึกและค่าความแม่นยำในเชิงฮาร์โมนิค (harmonic) เหมาะสำหรับฐานข้อมูลสารสนเทศที่มีขนาดใหญ่มาก และมักจะไม่สามารถหาข้อมูลภาพที่ถูกต้องทั้งหมดมีอยู่เท่าใด ทำให้ต้องทำการประมาณโดยใช้การสุ่มตัวอย่าง (sampling) ตามหลักทางสถิติหรือด้วยวิธีอื่นด้วย โดยทั่วไปจะเป็นการหาค่า F-measure ซึ่งแสดงสูตรได้ดังนี้

$$F = \frac{2(\text{Pr} \cdot \text{Re})}{(\text{Pr} + \text{Re})}$$

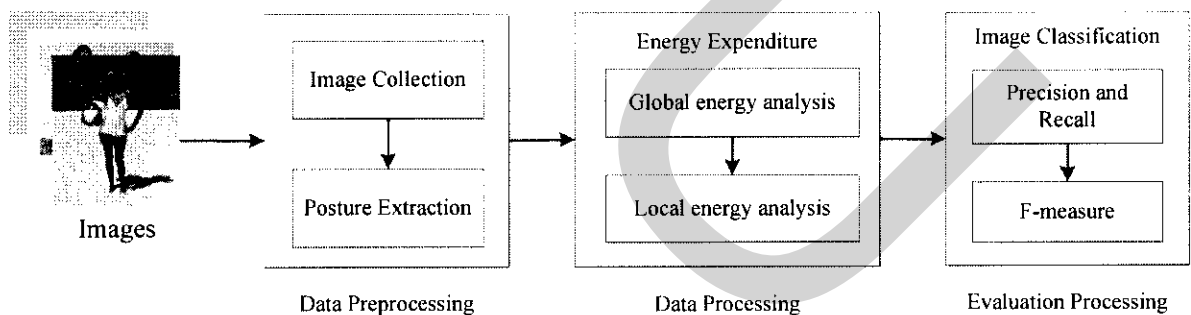
บทที่ 3

ขั้นตอนวิธีการดำเนินงานวิจัย

ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยนี้ได้มีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยออกเป็น 3 ส่วนหลักดังนี้

1. ขั้นตอนการเตรียมข้อมูล (data preprocessing) ประกอบด้วย การจัดเก็บภาพ (image collection) เป็นการคัดเลือกฐานข้อมูลภาพที่เหมาะสมกับการทดลอง และทำการสกัดท่าทางออกจากภาพ (posture extraction) เป็นการแม็พภาพเพื่อจัดเก็บข้อมูลเป็นพีเจอร์
2. ขั้นตอนการประมวลผล (data processing) เป็นการประมวลผลของโมเดลที่นำเสนอคือ Energy expenditure ที่ทำการแปลงค่าพีเจอร์เป็นค่าพลังงานเพื่อบอกความหมายของท่าทางบนภาพ
3. ขั้นตอนการวัดประสิทธิภาพ (evaluation processing) ทำการจำแนกท่าทางของภาพให้อยู่ในกลุ่มที่มีการกำหนดไว้

ขั้นตอนทั้งหมดถูกแสดงไว้ในภาพที่ 3.1

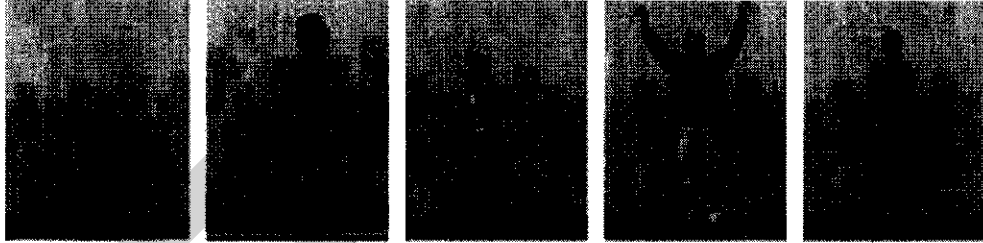


ภาพที่ 3.1 ขั้นตอนการประมวลผล

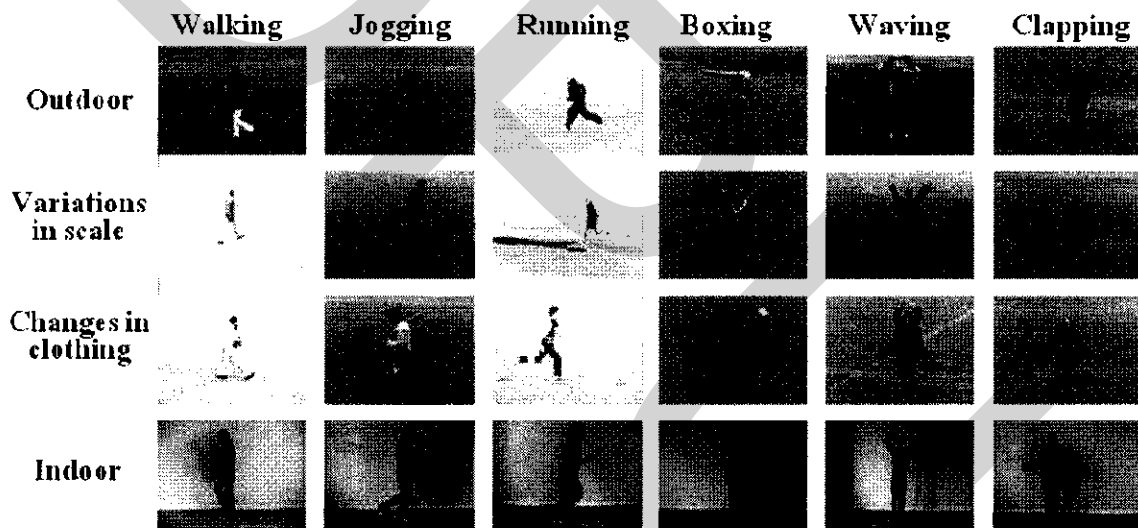
3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการทำการแยกและคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ สำหรับภาพบางภาพจะไม่นำเข้ามา

ทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวม ภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้



ภาพที่ 3.2 ภาพตัวอย่างชุดข้อมูลของ Weizmann [Lena Gorelick 2005]



ภาพที่ 3.3 ภาพตัวอย่างชุดข้อมูลของ KTH [Schuldt C. 2004]

3.1.1 ฐานข้อมูลรูปภาพ

การหาแหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ฐานข้อมูลภาพที่สมบูรณ์เพียงพอที่จะสามารถนำมาใช้งานได้โดยไม่มีผลข้างเคียงต่อกระบวนการที่นำเสนอ ดังนั้น แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพ

ได้ เช่น Fotosearch stock² The Cobis Stock³ The Corel Corporation⁴ เป็นต้น เป็นคลังข้อมูลภาพแต่อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอเนื่องจาก ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดเจนหรือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ภาพบางภาพอาจไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถหาความหมายภาพได้ ทำให้ต้องมีการคัดเลือก ใช้เฉพาะบางภาพเพื่อเป็นข้อมูลในการทดลองในการทดสอบวิธีการที่นำเสนอ



ภาพที่ 3.4 ตัวอย่างภาพข้อมูลทำทางเงาจากการ threshold [Nazli Ikizler-Cinbis 2009]

สำหรับชุดข้อมูลที่จะนำมาใช้เป็นข้อมูลภาพพื้นฐานการรู้จำท่าทางของงานวิจัยนี้ที่นำมาใช้จะมีภาพพื้นฐานหลังเป็นยูนิฟอร์มที่นำมาจาก Weizmann [Lena Gorelick 2005] โดยมีการแบ่งภาพเป็นกลุ่มย่อยดังนี้ walk, run, jump, gallop sideways, bend, one-hand wave, two-hands wave, jump in place, jumping Jack และ skip ดังภาพตัวอย่างในภาพที่ 3.2 และ ชุดข้อมูล KTH [Schuldt C. 2004] โดยมีการแบ่งภาพเป็นกลุ่มย่อยดังนี้ walking, jogging, running, boxing, hand-waving และ hand-clapping ดังภาพตัวอย่างในภาพที่ 3.3 โดยอยู่ภายใต้ของพื้นหลังที่มีลักษณะต่างกัน 4 แบบดังนี้ indoor, outdoor, clothing และ การเปลี่ยนแปลงขนาดของวัตถุภายในภาพ (variations in scale)

3.1.2 การสกัดพีเจอร์ท่าทางมนุษย์

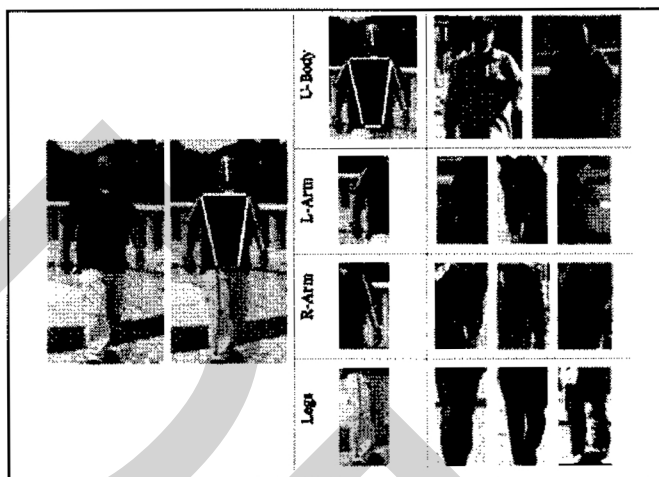
สิ่งที่สำคัญถัดจากการเลือกใช้ฐานข้อมูลสำหรับกระบวนการรู้จำท่าทางมนุษย์ คือการสกัดพีเจอร์บนภาพเพื่อนำมาใช้ในการประมวลผลข้อมูลต่อไป พีเจอร์บนภาพที่สนใจและต้องการนำมาใช้ในการทดลองคือท่าทางมนุษย์ วิธีการสกัดพีเจอร์เพื่อนำข้อมูลท่าทางมนุษย์มาใช้ในการทดลองมีหลายวิธี เช่น R. Nazli et al., [Nazli Ikizler-Cinbis 2009] ได้ใช้วิธีดึงท่าทางจากภาพหนึ่งด้วยการใช้เอ็ดจ์และพีเจอร์จากบริเวณรอบร่างกาย เพื่อใช้ในการสร้างโมเดล โดยการ

² Fotosearch Stock: <http://www.fotosearch.com>

³ The Corbis Stock: <http://pro.corbis.com>

⁴ The Corel Corporation: <http://www.corel.com/>

ทำงานของโมเดลนี้จะถูกแทนด้วยโครงสร้างของต้นไม้ เพื่อให้เป็นโมเดลดีฟอร์มในส่วนบริเวณโดยรอบในการสร้างพีเจอร์ร่างกายดังแสดงในภาพที่ 3.4 Moez Baccouche [Moez Baccouche 2011] ใช้กระบวนการรู้จำท่าทางด้วย 2 ขั้นตอนจาก Neural-based deep model ที่มีการสร้างจาก Convolutional Neural Networks เป็น 3D case เพื่อใช้ในการเรียนรู้พีเจอร์แบบอัตโนมัติ

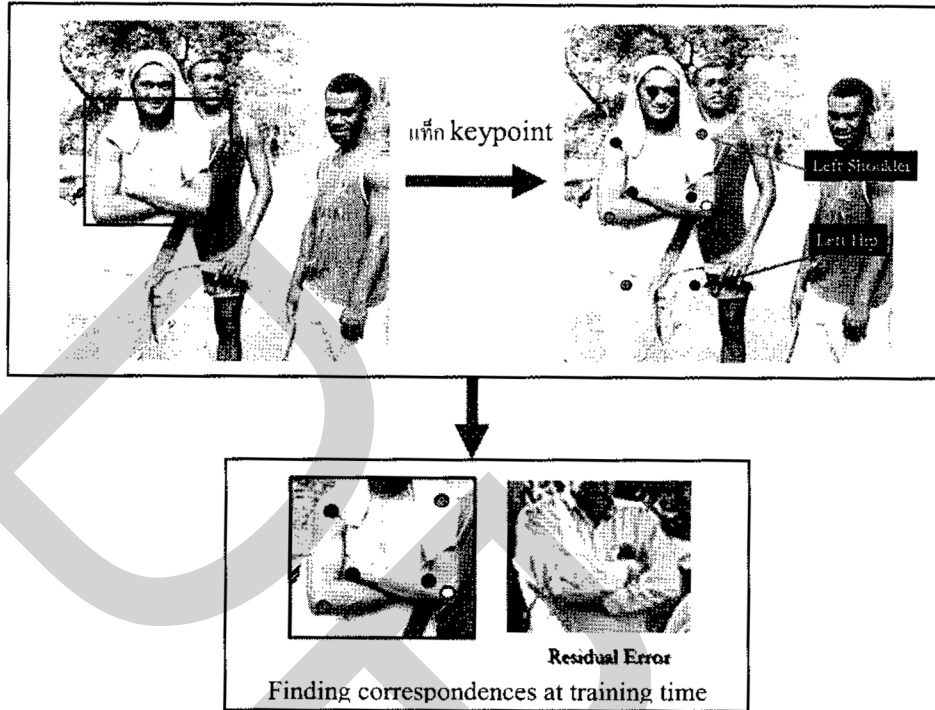


ภาพที่ 3.5 ตัวอย่างภาพข้อมูลที่มีความีเซตของ patch ใกล้เคียงกัน [Weilong Yang 2010]

Weilong Yang et al., [Weilong Yang 2010] ใช้วิธีการเลือกภาพที่มีการแสดงออกของท่าทางต่างๆ ที่อาจจะไม่ได้อยู่ในท่าทางที่เหมาะสมก็สามารถนำมาใช้ในการประมวลผล เพื่อรู้จำท่าทางด้วยวิธี Coarse exemplar based ดังแสดงในภาพที่ 3.5 คือวิธีการ “poselet” [L. Bourdev 2009] เป็นการอ้างอิงถึงเซตของ patch โดยในส่วนของเซตนี้จะมีความหมายที่เหมือนกันเมื่อมีท่าทางเดียวกัน หรือกล่าวอีกนัยหนึ่งว่าเมื่อมีการกระทำในกลุ่มเดียวกันจะมีค่าภายในเซต patch แบบเดียวกันดังแสดงในภาพที่ 3.6 จะเห็นว่าภาพพื้นหลังไม่คงที่แต่ในส่วนของร่างกายจะมีการแสดงที่คล้ายกัน โดยมีการแยกวิเคราะห์เป็น 4 ส่วนดังนี้ U-body L-arm R-arm และ Legs

จากขั้นตอนการสกัดพีเจอร์ที่กล่าวมา จะนำมาประยุกต์ใช้ภายในงานวิจัยนี้เพื่อเป็นแนวทางในการความหมายของท่าทางการเคลื่อนไหวของร่างกาย ในรูปลักษณะของการแท็ก keypoint บนโครงสร้างร่างกายตามข้อต่อบนภาพ [L. Bourdev 2009] เพื่อนำมาใช้เป็นข้อมูลเบื้องต้นในการทดลอง ข้อมูลที่ได้จากภาพจะเป็นการกำหนดค่าองศาของแต่ละตำแหน่งข้อต่อ (joint position) บนร่างกาย ทั้งหมดรวม 11 จุด ดังนี้ ส่วนของร่างกายท่อนบน (upper body) ด้านซ้ายและขวา ประกอบด้วย หลัง (torso) แขนส่วนบน (upper arms) แขนส่วนล่าง (forearms) และส่วนด้านล่างของร่างกายด้านซ้ายและขวา ประกอบด้วย ต้นขา (thighs) และ ขา

ส่วนล่าง (legs) ดังแสดงภาพตัวอย่างในภาพที่ 3.7 แสดงการแม็พส่วนข้อต่อบนร่างกายตามจุดต่าง ๆ เพื่อวัดค่าตำแหน่งองศาทั้ง 11 ข้อต่อค่าที่ได้จะถูกจัดเก็บลงในตัวแปรเพื่อเป็นฟีเจอร์ในการคำนวณค่าพลังงานในส่วนถัดไป



ภาพที่ 3.6 ขั้นตอนการ poselet [L. Bourdev 2009]

3.2 ขั้นตอนการประมวลผล

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลภาพ และภาพข้อมูลได้ถูกคัดเลือกเฉพาะภาพที่เหมาะสม และทำการแม็พโครงสร้างร่างกายเพื่อทำสกัดข้อมูลจากท่าทางการเคลื่อนไหวของร่างกายมนุษย์ จะได้มาซึ่งตำแหน่งองศาของข้อต่อทั้งหมด และทำการเก็บเป็นข้อมูลฟีเจอร์เข้า ใช้สำหรับทำการประมวลผลตามโมเดลของการใช้พลังงาน (Energy Expenditure) ที่นำเสนอ

การวิเคราะห์โมเดลการใช้พลังงานของร่างกายมนุษย์เพื่อหาความหมายของภาพโดยรวมที่ได้มาจากท่าทางการเคลื่อนไหว โมเดลนี้ได้ดัดแปลงจากทฤษฎี Physical Anthropology [MMWR 2001] เป็นทฤษฎีพื้นฐานทางชีววิทยามนุษย์ที่ศึกษาถึงความเข้มข้นในการใช้พลังงานในการเคลื่อนไหวจากท่าทางมนุษย์ด้วยหลักการทางชีววิทยาของร่างกายมนุษย์ ทางด้าน

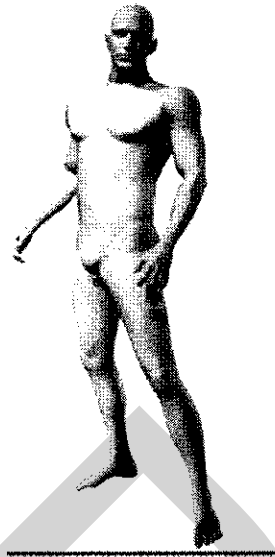
กลศาสตร์การเคลื่อนไหว [D. A. Winter 2005] ดังนั้นในงานวิจัยนี้ได้นำทฤษฎีมาประยุกต์ใช้บอกลักษณะความหมายของท่าทางมนุษย์เพื่อแปลเป็นความหมายของภาพ



ก. ท่าทางยืน ข. ท่าทางนั่ง ค. ท่าทางวิ่ง ง. ท่าทางกระโดด

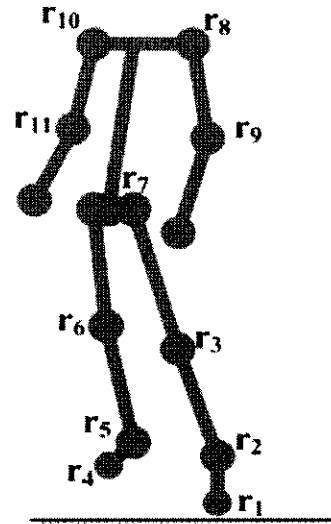
ภาพที่ 3.7 โครงร่างข้อต่อบนร่างกาย

โมเดลร่างกายมนุษย์ (human model) จะถูกแทนด้วยโมเดลสองมิติ stick figure ที่แทนถึงโครงสร้างร่างกายมนุษย์โดยเฉพาะในส่วนของตำแหน่งข้อต่อที่เป็นจุดสนใจสำหรับการทดลองนี้ ดังแสดงในภาพที่ 3.8 ก. แสดงถึงโมเดลร่างกายมนุษย์ และภาพที่ 3.8 ข. แสดงถึงโครงร่างข้อต่อบนโมเดลสองมิติ stick figure ที่ประกอบด้วย 11 จุดข้อต่อเป็นตำแหน่งอ้างอิง (reference point) ดังนี้ ส่วนของร่างกายท่อนบน (upper body) ด้านซ้ายและขวา ประกอบด้วย หลัง (torso) แขนส่วนบน (upper arms) แขนส่วนล่าง (forearms) และส่วนด้านล่างของร่างกายด้านซ้ายและขวา ประกอบด้วย ต้นขา (thighs) และ ขาส่วนล่าง (legs) แทนด้วย r_i เมื่อ i คือ ตัวเลขจำนวนเต็มที่มีค่าตั้งแต่ 1 ถึง 11 ทำการตั้งแกน x และ y เพื่อวัดค่าตำแหน่งองศาทั้ง 11 ข้อต่อสามารถวัดได้ดังแสดงในภาพที่ 3.9 แสดงตัวอย่างการวัดค่าองศาจากตำแหน่งแกนบนข้อต่อ ค่าที่ได้จะถูกจัดเก็บลงในตัวแปร $d = \{\theta_1, \dots, \theta_i, \dots, \theta_v\}$ เมื่อ v คือ ตัวเลขจำนวนเต็มที่มีค่า 11 เพื่อเป็นพีเจอรในการคำนวณค่าพลังงาน โดยแบ่งการคำนวณบนร่างกายออกเป็น 2 ส่วนดังนี้



ก.

ก. โมเดลร่างกายมนุษย์



ข.

ข. โครงร่างข้อต่อและตำแหน่งอ้างอิง

ภาพที่ 3.8 โมเดลสองมิติ stick figure

3.2.1 การวิเคราะห์ค่าพลังงานรวม

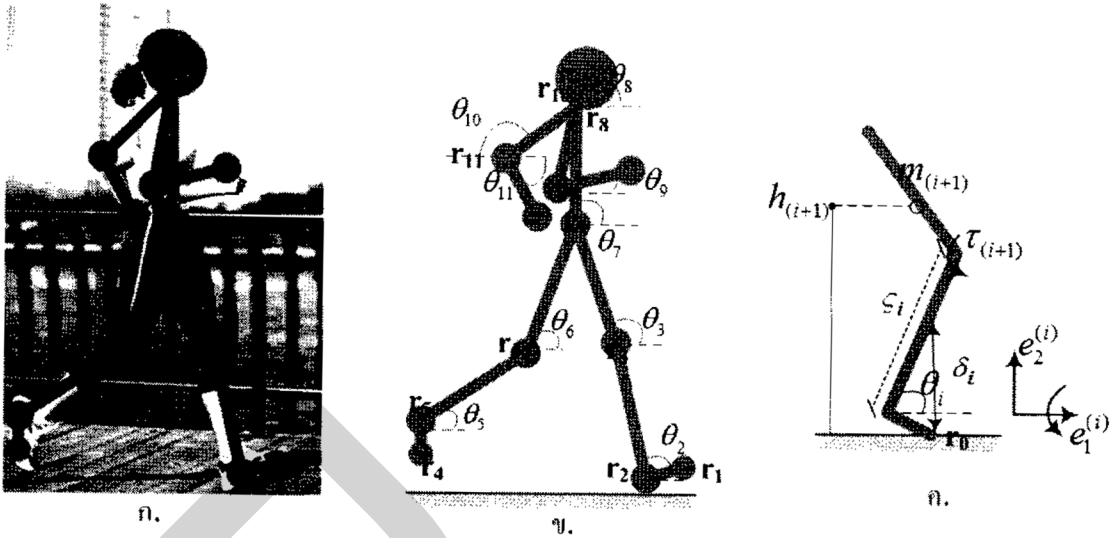
การวิเคราะห์ค่าพลังงานรวม (Global energy analysis: GE) เป็นการวิเคราะห์ค่าพลังงานโดยรวมของร่างกายที่มาจากแรงโน้มถ่วง (forces of gravity: g) [D. A. Winter 2005] ด้วยค่าคงที่ 9.8 m/sc . แสดงในภาพที่ 3.9 ข. จะทำการคำนวณค่าที่ได้รับมาจากข้อมูลเข้า d , $d = \{\theta_1, \dots, \theta_i, \dots, \theta_v\}$ เมื่อ v คือ จำนวนจุดอ้างอิงบนตำแหน่งข้อต่อที่มีค่าคงที่เป็น 11 และเมื่อ i คือ ตัวเลขจำนวนเต็มตามตำแหน่งอ้างอิงที่มีค่าตั้งแต่ 1 ถึง 11 ดังนั้นในจุดอ้างอิงที่ i แทนได้ด้วย r_i สามารถคำนวณค่าของ GE ได้ดังสมการต่อไปนี้

$$GE = \frac{\sum_{i=1}^n m_i \cdot g \cdot h_i}{\sum_{i=1}^n m_i \cdot g},$$

เมื่อ h_i แทนค่าความสูงของร่างกายในตำแหน่งที่ r_i , l_i แทนค่าความยาวของร่างกายในตำแหน่งที่ r_i , m_i แทนค่ามวลของร่างกายในตำแหน่งที่ r_i โดยที่ค่าของ h_i ได้มาจากผลรวมของความยาวส่วนของร่างกายตาม r_i ที่มีความสัมพันธ์กับจุดอ้างอิงหลักบนร่างกายคือ ตำแหน่งที่ 0 (r_0)

$$h_i = \sum_{j=0}^{j=i} l_j \cdot \delta_j.$$

เมื่อ δ_j แทนเมตริกซ์จุดหมุนบนแกน x ($\hat{e}_1^{(i)}$) และแกน y ($\hat{e}_2^{(i)}$) แสดงในภาพที่ 3.9 ค. สามารถเขียนเป็นสมการได้ดังนี้



ก. การแม็พเป็นโครงร่างข้อต่อ ข. วัตถุประสงค์ข้อต่อบนโครงร่างข้อต่อ ค. การหมุนแกนวัดค่า
 ภาพที่ 3.9 ตัวอย่างภาพวัดค่าตำแหน่งองศาบนข้อต่อร่างกาย

$$\delta_i = \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix} \cdot \begin{bmatrix} \hat{e}_1^{(i)} \\ \hat{e}_2^{(i)} \end{bmatrix}$$

โดยที่

$$\begin{aligned} \hat{e}_1^{(i)} &= \cos \theta_i \cdot e_1^{(i-1)} + \sin \theta_i \cdot e_2^{(i-1)}, \\ \hat{e}_2^{(i)} &= \cos \theta_i \cdot e_2^{(i-1)} - \sin \theta_i \cdot e_1^{(i-1)}. \end{aligned}$$

เมื่อ θ_i แทนค่าองศาจากตำแหน่งแกนบนข้อต่อที่ r_i

3.2.2 การวิเคราะห์ค่าพลังงานส่วนย่อย

การวิเคราะห์ค่าพลังงานส่วนย่อย (Local energy analysis) เป็นการหาค่าพลังงานตามส่วนของร่างกายในแต่ละข้อต่อ r_i ที่มีความสัมพันธ์กับจุดอ้างอิงหลักบนร่างกาย จะประกอบด้วยค่าพลังงานศักยภาพ (potential energy : PE), $PE = \{PE_1, \dots, PE_i, \dots, PE_v\}$ และค่าพลังงานแรงบิด (torque : Γ), $\Gamma = \{\Gamma_1, \dots, \Gamma_i, \dots, \Gamma_v\}$ แสดงสมการได้ดังนี้

$$PE_i = l_i \cdot h_i,$$

เมื่อ PE_i แทนค่าพลังงานศักยภาพที่ตำแหน่ง r_i ที่มีความสัมพันธ์กับแรงโน้มถ่วง, h_i แทนค่าความสูงของร่างกายในตำแหน่งที่ r_i และ l_i แทนค่าความยาวของร่างกายในตำแหน่งที่ r_i

$$\Gamma_i = F_i \times \lambda_i,$$

เมื่อ Γ_i แทนค่าพลังงานแรงบิดที่ตำแหน่ง r_i

$$F_i = m_i \cdot g,$$

เมื่อ F_i แทนแรงที่เกิดขึ้นจากมวลน้ำหนัก ที่มีความสัมพันธ์กับแรงโน้มถ่วงที่ตำแหน่ง r_i , m_i ค่ามวลส่วนร่างกายที่ตำแหน่ง r_i ,

$$\lambda_i = \sum_{j=1}^{j=i} P^{(j,j-1)},$$

เมื่อ λ_i แทนค่าระยะทางรวมที่เกิดขึ้นระหว่างตำแหน่งจุดอ้างอิง และ $P^{(i,j-1)}$ จะได้จากการคำนวณของเมตริกซ์จุดหมุนของแกน x ($\hat{e}_1^{(i)}$) และแกน y ($\hat{e}_2^{(i)}$)

$$\text{เมื่อ } P^{(i,j-1)} = l_i \begin{bmatrix} \cos \theta_i & \sin \theta_i \\ -\sin \theta_i & \cos \theta_i \end{bmatrix} \cdot \begin{bmatrix} \hat{e}_1^{(i-1)} \\ \hat{e}_2^{(i-1)} \end{bmatrix}.$$

ดังนั้นค่าพลังงานที่เกิดขึ้นทั้งหมดจะถูกจัดเก็บลงบน $\kappa^{(1)} = \{k_1^{(1)}, \dots, k_j^{(1)}, \dots, k_c^{(1)}\}$

3.3 ขั้นตอนการวัดประสิทธิภาพ

ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ จะนำเอาผลการจำแนกกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่นำเสนอข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปของค่าความถูกต้อง (accuracy) และ F-measure ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปที่สามารถเข้าใจได้ง่าย เช่น ตาราง หรือกราฟซึ่งในงานวิจัยนี้ได้แสดงในรูปของตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) สามารถอ่านเพิ่มเติมได้ในบทที่ 2

บทที่ 4

ผลการทดลอง

ในการทำวิจัยครั้งนี้เพื่อหาความหมายของท่าทางมนุษย์จากภาพ ด้วยการแมปโครงสร้างร่างกาย ในข้อต่อที่สำคัญใช้เป็นตำแหน่งอ้างอิง เพื่อทำการเก็บเป็นข้อมูลพีเจอร์รี่ใช้สำหรับการประมวลผล ตามโมเดลของการใช้พลังงาน (Energy Expenditure) เพื่อหาความหมายของท่าทางการเคลื่อนไหว ที่ประกอบด้วย การวิเคราะห์ค่าพลังงานรวม (Global energy analysis) และการวิเคราะห์ค่าพลังงาน ส่วนย่อย (Local energy analysis) ทั้งสองส่วนเป็นการหาค่าพลังงานทั้งร่างกายและแบ่งตามส่วน ข้อต่อของร่างกายเพื่อใช้ในการบอกความหมายของท่าทางที่เกิดขึ้นบนภาพและทำเปรียบเทียบความ เหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาท เทียมแบบหลายชั้น แผนผังการจัดระบบตัวเอง เครือข่ายแบบเบย์ และซัพพอร์ตเวกเตอร์แมชชีน แสดงผลการทดลองดังนี้

4.1 การกำหนดข้อมูลภาพ

ข้อมูลสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก The Cobis Stock และภาพ พื้นฐานหลังเป็นยูนิฟอร์มที่นำมาจาก Weizmann [Lena Gorelick 2005] และชุดข้อมูล KTH [Schuldt C. 2004] ซึ่งได้มีการคัดเลือกภาพสำหรับการทดลองให้อยู่ในหมวดหมู่ในวงจำกัดเพียง 5 กลุ่มเพื่อทดสอบการจำแนกความหมายของข้อมูลภาพที่มีมนุษย์เป็นส่วนหลักในภาพ ประกอบด้วย กลุ่มท่าทางของมนุษย์ดังนี้ ยืน (standing), วิ่ง (running), นั่ง (sitting), กระโดด (jumping), และ นอน (lying) ซึ่งภาพที่นำมาถูกจำแนกความเหมือนกันของความหมายท่าทางการเคลื่อนไหวภาพ ด้วยการให้ค่าของพลังงานในแต่ละข้อต่อที่ทำการคำนวณได้จากภาพ และได้ทำการทดลองจำแนก ด้วยโครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines) ดังนั้นในการทดลองจะมีการแสดงผลสัฟต์ด้วยตาราง เชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) ได้แสดงผลการทดลองดังนี้

4.2 ผลการทดลอง

จากการทำการทดลอง เพื่อจำแนกความหมายของข้อมูลภาพออกโดยในงานวิจัยนี้จะแบ่งออกเป็น 5 กลุ่มประกอบด้วยกลุ่มของ กลุ่มที่ 1 แสดงถึงภาพที่มีคนยืน (standing) กลุ่มที่ 2 ภาพที่มีคนวิ่ง (running), กลุ่มที่ 3 ภาพที่มีคนนั่ง (sitting), กลุ่มที่ 4 ภาพที่มีคนกระโดด (jumping), และกลุ่มที่ 5 ภาพที่มีคนนอน (lying) โดยที่ในแต่ละกลุ่มจะต้องมีคนเป็นส่วนหลักที่เห็นอย่างเด่นชัด อาจจะมีภาพพื้นหลังที่ทำการแบ่งออกเป็น 2 ส่วนคือ

- ภาพพื้นหลังที่เป็นภายนอก (outdoor) มีสภาพแวดล้อมเป็นสนามหญ้า สนามเด็กเล่นหรือถนน เพื่อใช้ในการพักผ่อน ทำกิจกรรมอื่นๆ ภาพเมือง (city) ที่ประกอบด้วย ถนน อาคาร ภาพพื้นหลังที่เป็นสนามกีฬา (stadium) ที่ประกอบด้วย สนามแข่งขันต่างๆ ที่มีอุปกรณ์กีฬาประกอบ ภาพพื้นหลังที่เป็นทะเล (beach) ประกอบด้วยส่วนต่างๆ เช่น ท้องฟ้า ทะเล พื้นน้ำ พื้นทราย หรือ ต้นไม้
- ภาพพื้นหลังที่ภายใน (indoor) มีสภาพแวดล้อมเป็นบ้าน สถานที่ทำงาน อาจมีโซฟา เครื่องเฟอร์นิเจอร์ หรือห้องครัว หรือมีภาพพื้นหลังเป็นสนามหญ้า (lawn) ภาพภายในห้อง (room) ประกอบเป็นพื้นหลังได้

โดยที่เครื่องมือการจำแนกทั้งหมด 4 วิธีด้วยกันคือ naïve-Bayes, Multilayer, SOM, และ SVM ผลของการจำแนกความหมายของภาพใน 5 กลุ่ม ดังภาพแสดงในตารางที่ 4.1 ถึง 4.4 ด้วยตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (%)

ตารางที่ 4.1 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี naïve-Bayes

Class	Confusion Matrix (%)					Performance (%)		
	standing	running	jumping	sitting	lying	Prec.	Recall	F ₁
standing	73	16	8	3	0	72.3	73.0	72.6
running	14	70	11	5	0	70.7	70.0	70.4
jumping	7	9	82	2	0	78.8	82.0	80.4
Sitting	7	4	1	75	13	72.1	75.0	73.5
Lying	0	0	2	19	79	85.9	79.0	82.3
Accuracy rate						75.8		

ตารางที่ 4.2 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี Multilayer

Class	Confusion Matrix (%)					Performance (%)		
	standing	running	jumping	sitting	lying	Prec.	Recall	F ₁
standing	76	14	8	2	0	75	76	76
running	8	78	9	5	0	78	78	78
jumping	5	2	89	2	2	81	89	85
Sitting	12	6	4	72	6	82	72	77
Lying	0	0	0	7	93	92	93	93
Accuracy rate						81.6		

ตารางที่ 4.3 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี SOM

Class	Confusion Matrix (%)					Performance (%)		
	standing	running	jumping	sitting	lying	Prec.	Recall	F ₁
standing	83	9	5	3	0	78	83	81
running	9	75	11	5	0	77	75	76
jumping	7	9	82	2	0	83	82	82
Sitting	7	4	1	81	7	84	81	82
Lying	0	0	0	6	94	93	94	94
Accuracy rate						83.0		

ตารางที่ 4.4 ผลลัพธ์การจำแนกความหมายภาพด้วยวิธี SVM

Class	Confusion Matrix (%)					Performance (%)		
	walking	running	jumping	sitting	lying	Prec.	Recall	F ₁
walking	77	13	9	1	0	79	77	78
running	11	82	5	2	0	80	82	81
jumping	7	5	88	0	0	86	88	87
sitting	2	3	0	89	6	92	89	90
lying	0	0	0	5	95	94	95	95
Accuracy rate						86.2		

จากผลการทดลองได้แสดงในตารางที่ 4.1 - 4.4 เป็นตารางเชื่อมโยงปฏิสัมพันธ์ความสับสนที่ประกอบด้วย 5 กลุ่มที่ประกอบด้วย ดังนี้ ยืน (standing), วิ่ง (running), นั่ง (sitting), กระโดด (jumping), และ นอน (lying) จะเห็นว่ากลุ่มท่าทาง running ได้ค่าความถูกต้องเพียง 70% กลุ่มท่าทาง jumping ได้ค่าความถูกต้องถึง 82% ด้วยวิธีการจำแนกแบบ naïve-Bayes แต่ได้ค่าเฉลี่ยของค่าความถูกต้องถึง 75.8% และการจำแนกด้วยวิธีการ Multilayer แต่สำหรับกลุ่มภาพที่มีท่าทาง running ได้ค่าความถูกต้องถึง 78% และกลุ่มท่าทาง lying จะมีค่าความถูกต้องมากถึง 93% และได้ค่าเฉลี่ยของค่าความถูกต้องถึง 81.6% และ สำหรับการจำแนกด้วย SOM ซึ่งได้ค่าเฉลี่ยความถูกต้องถึง 94% ซึ่งมากกว่าการจำแนกด้วย naïve-Bayes 75.8% และ Multilayer 81.6% แต่อย่างไรก็ตามการจำแนกในกลุ่มภาพ lying ด้วยวิธี SOM จะได้ค่าความถูกต้องมากถึง 94% จะสังเกตว่าในการจำแนกด้วยวิธี SVM การจำแนกในกลุ่มของท่าทางการยืน วิ่ง และกระโดด นั้นจะมีค่าความถูกต้องที่น้อยกว่า ท่าทางนั่ง sitting ที่ได้ถึง 89 % และท่าทางนอน lying ได้ถึง 95% มีค่าเฉลี่ยของค่าความถูกต้องถึง 86.2% แต่อย่างใดก็ตามท่าทาง lying เป็นท่าทางที่สามารถจำแนกได้ดีมาก จะเห็นว่า Multilayer 93% SOM 94% เนื่องจากการใช้ค่าพลังงานของท่าทางในส่วนของการวิเคราะห์ค่าพลังงานรวม สามารถที่จะแยกแยะท่าทาง lying ออกจากท่าทางการยืนที่มี standing running และ jumping ได้อย่างชัดเจน แต่ในส่วนของท่าทาง sitting นั้นค่าพลังงานที่ได้จะอยู่ระหว่างท่าทางยืนและท่าทางนอนจึงไม่สามารถจำแนกได้อย่างชัดเจน จึงจำเป็นต้องมีการใช้การวิเคราะห์ค่าพลังงานส่วนย่อย เข้ามาใช้เป็นการหาค่าพลังงานทั้งร่างกายและแบ่งตามส่วนข้อต่อของร่างกายเพื่อใช้ในการบอกความหมาย ดังนั้นในการจำแนกท่าทางการยืนที่มี standing running และ jumping จะสามารถจำแนกค่อนข้างยากเนื่องจากจะต้องใช้ค่าของพลังงานส่วนย่อย ตามข้อต่อต่างๆ จึงทำให้ค่าความถูกต้องค่อนข้างต่ำกว่าท่าทางนอน การจำแนกด้วยวิธี SOM ท่าทาง standing ได้ 83% มีค่าความแม่นยำ 78% ค่าความระลึกลับ 83% SVM ท่าทาง jumping ได้ถึง 88% มีค่าความแม่นยำ 86% ค่าความระลึกลับ 88% การจำแนกท่าทางของมนุษย์บนภาพที่มีสภาพแวดล้อมแตกต่างกันสามารถจำแนกได้ด้วยค่าของพลังงาน 2 ส่วนคือ ค่าพลังงานรวม และ ค่าพลังงานส่วนย่อย โดยที่ใช้วิธีการจำแนกด้วย SVM จะได้ค่าความถูกต้องถึง 86.2% SOM 83% Multilayer 81.6% และ naïve-Bayes 75.8%



ก. ยืน (standing)



ข. วิ่ง (running)



ค. กระโดด (jumping)



ง. นั่ง (sitting)



จ. นอน (lying)

ภาพที่ 4.1 ตัวอย่างผลลัพธ์ของการจำแนกความหมายภาพ

เมื่อกล่าวโดยสรุป สำหรับพีเจอร์ที่ถูกสกัดมาจากท่าทางการเคลื่อนไหวเป็นมุมค่าองศาที่ทำมุมกับแกนหมุนเพื่อนำมาคิดค่าพลังงานของท่าทางการเคลื่อนไหวจะได้ค่าความถูกต้องเฉลี่ยรวมที่มีค่ามากที่สุด ถึง 86.2% เมื่อใช้การจำแนกด้วยวิธี SVM ในขณะที่ naïve-Bayes Multilayer และ SOM มีค่าความถูกต้องเฉลี่ยรวมเพียง 75.8% 81.6% และ 83% ตามลำดับ ดังนั้นจะเห็นว่าการใช้ความสัมพันธ์ของค่าพลังงานที่มีความสัมพันธ์ทางสรีระร่างกายมาใช้ในการพิจารณาความหมายของท่าทาง สามารถจำแนกได้แต่บางท่าทางที่มีความหมายกำกวม หรือมีการบิดบังบางส่วนจากร่างกาย การคำนวณด้วยวิธีนี้จะค่อนข้างลำบาก และถ้าท่าทางที่มีลักษณะที่ใกล้เคียงกันมากค่าพลังงานจะไม่สามารถจำแนกได้ เช่น การนั่งแล้วยกขาสองข้างขึ้น ทำให้ค่าพลังงานบางส่วนของท่านขาล่างจะมีค่าพลังงานที่ผิดแปลกทำให้การจำแนกได้ค่อนข้างยาก เป็นต้น ภาพผลลัพธ์ของการจำแนกในภาพที่ 4.1 แสดงกลุ่มท่าทางดังนี้ ยืน (standing), วิ่ง (running), นั่ง (sitting), กระโดด (jumping), และ นอน (lying)

บทที่ 5

สรุปผลการทดลอง

ในบทนี้ได้ทำการสรุปผลการทดลอง และข้อเสนอแนะ รวมทั้งข้อผิดพลาดที่เกิดขึ้นจากการทดลอง รวมไปถึงสิ่งที่ควรปรับปรุงเพิ่มเติม เพื่อแก้ไขข้อผิดพลาดและแนวทางการทำวิจัยต่อในเรื่องของ semantic image ซึ่งเป็นงานวิจัยที่ปัจจุบันได้มีนักวิจัยให้ความสนใจอย่างแพร่หลายสามารถนำไปประยุกต์ใช้ได้หลายด้าน เช่น การแพทย์ การขนส่งคมนาคม การรักษาความปลอดภัย หรืออุตสาหกรรม เป็นต้น

5.1 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้เสนอหัวข้อวิจัยทางการประมวลผลภาพ ในส่วนของการจำแนกท่าทางการเคลื่อนไหวของมนุษย์บนภาพ ด้วยการแม็พโครงสร้างร่างกายในข้อต่อที่สำคัญใช้เป็นตำแหน่งอ้างอิงเพื่อทำการเก็บเป็นข้อมูลพีเจอร์ีใช้สำหรับทำการประมวลผล ตามโมเดลของการใช้พลังงาน (Energy Expenditure) เพื่อหาความหมายของท่าทางการเคลื่อนไหวที่ประกอบด้วย การวิเคราะห์ค่าพลังงานรวม (Global energy analysis) และการวิเคราะห์ค่าพลังงานส่วนย่อย (Local energy analysis) ทั้งสองส่วนเป็นการหาค่าพลังงานทั้งร่างกายและแบ่งตามส่วนข้อต่อของร่างกายเพื่อใช้ในการบอกความหมายของท่าทางที่เกิดขึ้นบนภาพเพื่อให้ได้ความหมายของภาพข้อมูลสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก The Cobis Stock และภาพพื้นฐานหลังเป็นยูนิฟอร์มที่นำมาจาก Weizmann [Lena Gorelick 2005] และชุดข้อมูล KTH [Schuldt C. 2004] ซึ่งได้มีการคัดเลือกข้อมูลภาพที่มีมนุษย์เป็นส่วนหลักในภาพ เพื่อทดสอบการจำแนกความหมายโดยทำการแบ่งเป็นกลุ่มท่าทางสำคัญหลักดังนี้ ยืน (standing), วิ่ง (running), นั่ง (sitting), กระโดด (jumping), และนอน (lying) ซึ่งภาพที่นำมาถูกจำแนกความเหมือนกันของความหมายท่าทางการเคลื่อนไหวภาพด้วยการใช้ค่าของพลังงานในแต่ละข้อต่อที่ทำการคำนวณได้จากภาพ ดังแสดงตัวอย่างของการสกัดพีเจอร์ีและแม็พโครงสร้างได้ในภาพที่ 5.1



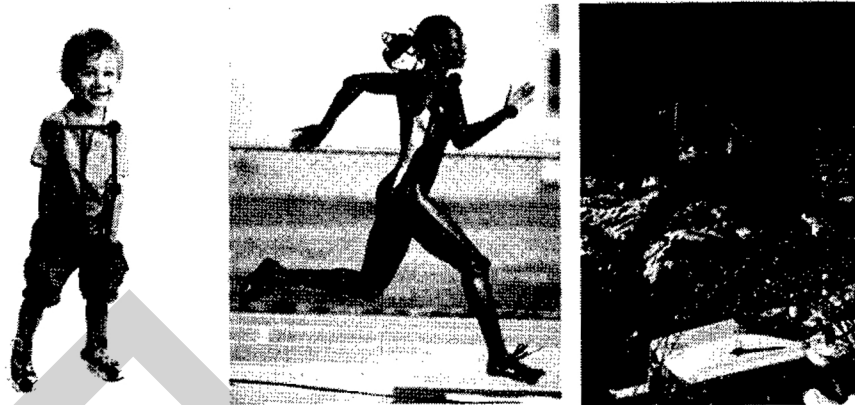
ก.นั่ง (sitting)



ข. กระโดด (jumping)

ภาพที่ 5.1 ตัวอย่างการแม็พโครงร่างข้อต่อบนร่างกาย

จากภาพที่ 5.1 แสดงตัวอย่างการแม็พโครงร่างข้อต่อของกลุ่มภาพ 2 กลุ่มคือ นั่ง และ กระโดด จะเห็นว่าลักษณะตำแหน่งการทำมุมของโครงร่างข้อต่อมีความแตกต่างกันอย่างเด่นชัดในสองกลุ่มนี้ กล่าวคือค่าพลังงานของกลุ่มนั่งจะได้ค่าพลังงานรวมและค่าพลังงานส่วนย่อยมีค่าที่ต่ำกว่ากลุ่มกระโดด แต่อย่างไรก็ตาม ในกลุ่มของท่าทางการนั่งในลักษณะของการนั่งบนพื้น และนั่งบนเก้าอี้ที่สูงกว่าพื้นการจำแนกค่าของพลังงานจะมีความแตกต่างกันค่อนข้างมากโดยเฉพาะค่าพลังงานรวมที่ให้ค่าแตกต่างกัน หมายถึง ค่าพลังงานรวมของการนั่งบนพื้นจะได้ค่าพลังงานที่ต่ำกว่าค่าพลังงานรวมที่ นั่งบนเก้าอี้ ถ้าท่าทางนั่งบนเก้าอี้ที่มีบางส่วนของร่างกายที่ยื่นขึ้นเหนือศีรษะ หรือกางแขน กางขา เกิดขึ้นจะทำการยิ่งเกิดค่าของพลังงานในส่วนย่อยมากกว่าปรกติทำให้การจำแนกบางส่วนผิดพลาดได้ ดังนั้นอาจจะต้องมีการแบ่งกลุ่มย่อยสำหรับท่าทางที่มีความแตกต่างเพิ่มมากขึ้น



ก. ยืน (standing) ข. วิ่ง (running) ค. กระโดด (jumping)

ภาพที่ 5.2 ตัวอย่างการแม็พโครงร่างข้อต่อบนร่างกาย

จากขั้นตอนการทดลองในบทที่ 4 สามารถสรุปได้ว่า การใช้ข้อมูลพลังงานที่ได้จากการสกัดมาจากท่าทางการเคลื่อนไหวเป็นมุมค่าองศาที่ทำมุมกับแกนหมุนเพื่อนำมาคิดค่าพลังงานของท่าทางการเคลื่อนไหวสามารถช่วยในการจำแนกความหมายของท่าทางมนุษย์ ได้ค่าความถูกต้องเฉลี่ยรวมที่มีค่ามากที่สุด ถึง 86.2% เมื่อใช้การจำแนกด้วยวิธี SVM ในขณะที่ naïve-Bayes Multilayer และ SOM มีค่าความถูกต้องเฉลี่ยรวมเพียง 75.8% 81.6% และ 83% ตามลำดับ จะเห็นว่าการใช้ความสัมพันธ์ของค่าพลังงานที่มีความสัมพันธ์ทางสรีระร่างกาย สามารถจำแนกได้อย่างชัดเจน ดังแสดงในภาพที่ 5.2 แสดงตัวอย่างในกลุ่มของ ท่าทางยืน ท่าทางวิ่ง และท่าทางกระโดด ซึ่งมีค่าพลังงานรวมของการยืนที่ใกล้เคียงกัน แต่จะมีความแตกต่างกันในส่วนของค่าพลังงานย่อยตามข้อต่อที่เกิดการหมุนของแกนที่ต่างกันเท่านั้น ส่วนของการกระโดดจะมีค่าพลังงานรวมที่เพิ่มมากขึ้นเนื่องจากเกิดความสูงของมวลรวมนั่นเอง

สำหรับการทดลองในครั้งนี้มีมีการจำแนกด้วยโครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และซัพพอร์ตเวกเตอร์แมชชีน (Support vector machines) ดังนั้นในการทดลองจะมีการแสดงผลด้วยตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix)

5.2 ข้อเสนอแนะ

ในกระบวนการจำแนกข้อมูลภาพออกเป็นกลุ่มย่อยที่มีความหมายในรูปแบบของความหมายภาพโดยรวมนั้นจะเป็นการพิจารณาภาพที่ซับซ้อนมากขึ้น นอกจากการค้นคืนข้อมูลภาพหรือการจำแนกภาพโดยทั่วไปที่ต้องการค้นคืนเพียง ความเหมือนกันของวัตถุที่เกิดขึ้นบนภาพเท่านั้น ในงานวิจัยส่วนใหญ่จึงเน้นไปที่การสกัดข้อมูลภาพในรูปแบบของการสกัดด้วยอัลกอริทึมที่แตกต่างกัน และค้นหาภาพเพียงความเหมือนกันของรูปร่างหรือลักษณะเฉพาะ หรือเพียงแต่วัตถุบนภาพ เท่านั้น ทั้งที่มีความหมายและไม่มี ความหมาย แต่อย่างไรก็ตาม เบื้องต้นของผลลัพธ์ที่ได้จากการค้นคืนหรือจำแนกภาพ คือความเหมือนกันทางกายภาพ เช่น รูปร่าง สี หรือชนิดของวัตถุ แต่ลักษณะการวิเคราะห์และพิจารณาของการเหมือนกันทางความหมายภาพ หรือ semantic นั้นจะมีลักษณะการวิเคราะห์ที่แตกต่างกันออกไป ในอีกรูปแบบหนึ่งซึ่งเป็นรูปแบบที่เกิดจากความคิดของมนุษย์ที่มีการแปลงความจากภาพ แต่อย่างไรก็ตามสิ่งที่ได้จากการทดลองนั้นยังคงมีข้อที่ต้องปรับปรุงเพิ่มอยู่อีกหลายส่วนด้วยกัน

สำหรับงานวิจัยนี้พยายามเน้นถึงภาพส่วนบุคคลที่มี มนุษย์แสดงออกถึงท่าทางเพื่อนำมาใช้ในการแปลความหมายของภาพเป็นส่วนหลัก แต่บางท่าทางที่มีความหมายกำกวม หรือมีการบิดบัง บางส่วนของร่างกาย การคำนวณด้วยวิธีนี้จะค่อนข้างลำบาก และถ้าท่าทางที่มีลักษณะที่ใกล้เคียงกันมากค่าพลังงานจะไม่สามารถจำแนกได้ เช่น การนั่งแล้วยกขาสองข้างขึ้น ทำให้ค่าพลังงานบางส่วนของท่อนขาล่างจะมีค่าพลังงานที่ผิดแปลกทำให้การจำแนกได้ค่อนข้างยาก เป็นต้น แต่อย่างไรก็ตามลักษณะของท่าทางและการจำแนกโดยรวมสามารถใช้โมเดลพลังงานได้เป็นอย่างดี

สิ่งควรจะมีการปรับปรุงเพิ่มเติมเพื่อให้การแปลความหมายของภาพได้ดียิ่งขึ้นอีก ก็คือการพิจารณารายละเอียดของวัตถุของภาพ ซึ่งคำหลักที่ถูกเลือกมาจากกลุ่มภาพใน LabelMe [A. Torralba 2010][B. C. Russell 2008] เพื่อเพิ่มความหลากหลายของกลุ่มคำศัพท์ควรจะมีการจัดกลุ่มของวัตถุที่เข้ามาทำการทดลองให้มีความหมายที่รัดกุมมากขึ้น และเมื่อภาพที่มีส่วนของสภาพแวดล้อมที่คล้ายคลึงกันมาก จะถูกจัดอยู่ในกลุ่มเดียวกันทั้งที่ วัตถุภายในภาพ และสามารถนำโมเดลค่าพลังงานมาประยุกต์ใช้ในการบอกท่าทางการกระทำในภาพเพื่อใช้ในการแปลความหมายของภาพโดยรวมได้ สำหรับภาพที่มี มนุษย์ ที่เป็นวัตถุเด่นในภาพ ถ้ามนุษย์มีรูปร่างลักษณะท่าทางที่แปลงเปลี่ยน เช่น ยืน นั่ง นอน หรือ กระโดด ทำให้การตีความหมายของภาพโดยรวมมีความหมายที่

แตกต่างกันออกไป และการพิจารณา อารมณ์ จากใบหน้าและท่าทางของการแสดงออกสัมพันธ์ควบคู่กันไป ซึ่งงานวิจัยในส่วนของอารมณ์มนุษย์ที่แสดงออกทางใบหน้า และท่าทางก็ยังคงเป็นงานวิจัยยังคงวิจัยกันอยู่มาก เพราะฉะนั้นในงานวิจัยที่นำเสนอนี้เป็นอีกแนวทางหนึ่งที่พยายามจะคิดค้นวิธีการที่จะหาความหมายที่เกิดขึ้นจากภาพ ในอีกมุมมองหนึ่งซึ่งยังคงต้องมีการพิจารณาและวิเคราะห์ปรับปรุงการทดลองต่อไป.

DUKU

บรรณานุกรม

- นัศพ์ชาณัณ ชินปัญช์ธนะ, เอกสารประกอบการสอนวิชา การประมวลผลภาพ (Image processing), มหาวิทยาลัยธุรกิจบัณฑิตย์, 2552.
- A. Branzan Albua, R. Bergevinb and S. Quirion, Generic temporal segmentation of cyclic human motion Pattern Recognition, Vol. 41, Issue 1, January, 2008, pp. 6–21.
- A. J. Lipton, H. Fujiyoshi, R.S. Patil “Moving Target Classification and Tracking from Real-time Video,” *IEEE Workshop on Applications of Computer Vision*, 1998.
- A. Torralba, B. C. Russell, J. Yuen, “LabelMe: Online Image Annotation and Applications,” Proceedings of the IEEE, Vol. 98, n. 8, August 2010, pp. 1467 – 1484.
- Apostol Paul Natsev , Atul Chadha , Basuki Soetarman , Jeffrey Scott Vitter, “CAMEL: Concept Annotated iMagE Libraries,” Storage and Retrieval for Media Databases Proceedings, vol. 4315, 2001.
- B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, “LabelMe: A Database and Web-based Tool for Image Annotation,” Intl. J. Computer Vision, vol 77, numbers 1-3, May, 2008, pp. 157–173.
- Bangpeng, Y., Fei-Fei, l., “Modeling Mutual Context of Object and Human Pose in Human-Object Interaction Activities,” In: CVPR. San Francisco, 2010.
- Ballan, L., Bertini, M., Del Bimbo, A., Seidenari, L. and Serra, G., “Effective Codebooks for human action categorization,” Computer Vision Workshops (ICCV Workshops), IEEE 12th International Conference , 506- 513, 2009.
- Benitez, A.B., and S.-F. Chang, “Semantic Knowledge Construction From Annotated Image Collections”, *International Conference On Multimedia & Expo (ICME-2002)*, Lausanne, Switzerland, Aug 26-29, 2002.
- C. Thureau and V. Hlaváč. Pose primitive based human action recognition in videos or still images. In CVPR, 2008.
- Carson, C., et al., “Blobworld: A system for regionbased image indexing and retrieval,” In Proc. Int. Conf. Visual Information System, 1999.
- Chester M, *Neural networks: a tutorial*, Prentice-Hall, Englewood Cliffs, 1993.
- D. A. Winter, *Biomechanics and motor control of human movement*, 3rd ed., John Wiley & Sons, Inc., 2005.

- D. Comaniciu, R. Visvanathan, P. Meer, "Kernel based object tracking". *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5), 2003, pp. 564–575.
- D. Comaniciu, V. Ramesh, P. Meer, "Real-time tracking of non rigid objects using mean shift". *Proceedings on Computer Vision and Pattern Recognition*, Hilton Head, vol. 1, 2000, pp. 142–149.
- F. Porikli, "Integral histogram a fast way to extract histograms in cartesian spaces," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- Flickner, M., et al., "Query by image and video content: The QBIC System," *IEEE Computer*, Vol. 28, No. 9, 1995.
- Galleguillos C., Belongie S., "Context Based Object Categorization: A Critical Survey," *Computer Vision and Image Understanding (CVIU)*, Vol. 114, 2010, pp. 712-722.
- Galleguillos C., McFee B., Belongie S., Lanckriet G., "From Region Similarity to Category Discovery," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, 2011.
- Galleguillos C., McFee B., Belongie S., Lanckriet G.R G., "Multi-Class Object Localization by Combining Local Contextual Interactions", *IEEE Conference in Computer Vision and Pattern Recognition (CVPR)*, 2010.
- Gupta, A., Kembhavi, A., Davis L.S., "Observing Human-Object Interactions: Using Spatial and Functional Compatibility for Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 1775-1789, 2009.
- Ikizler, N., Duygulu, P.I., "Histogram of oriented rectangles: A new pose descriptor for human action recognition," *IVC*. 27, 10, 1515–1526, 2009.
- Ikizler, N., R. Gokberk Cinbis, "Stan Sclaroff, Learning actions from the Web," *IEEE 12th International Conference on Computer Vision*, 995- 1002, 2009.
- James Z. Wang, Jia Li, Gio Wiederhold, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 23, No.9, 2001, pp. 947-963.
- Javed Ahmed, M. N. Jafri, Mubarak Shah, Muhammad Akbar, "Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking," *Machine Vision and Applications*, Vol. 19, Issue 1, January, 2008, pp. 1-25.
- Jia Li, James Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, 2003, pp. 1075-1088.

- Joo-Hwee Lim, Mulhem, P., Qi Tian, “Event-based home photo retrieval,” In Proceedings of Multimedia and Expo (ICME), July 2003.
- Kobus Barnard and David Forsyth, “Exploiting Image Semantics for Picture Libraries”, *The First ACM/IEEE-CS Joint Conference on Digital Libraries*, 2001.
- Kjellström, H., Romero, J., Martínez, D., Kragić, D., “Simultaneous Visual Recognition of Manipulation Actions and Manipulated Objects,” In: ECCV. Marseille, 2008.
- L. Bourdev and J. Malik. Poselets: Body part detectors training using 3d human pose annotations. In ICCV, 2009.
- Lazebnik S., Schmid, C., Ponce, J., “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” In: CVPR. New York, 2006.
- Lena Gorelick, Moshe Blank, Eli Shechtman, Michal Irani and Ronen Basri, Actions as Space-Time Shapes, Tenth IEEE International Conference on Computer Vision (ICCV), 2005. Li, L.J., Fei-Fei, L., “What, where and who? Classifying event by scene and object Recognition,” In: ICCV. Rio de Janeiro , 2007.
- Lippmann, R.P., “An Introduction to Computing with Neural Nets,” IEEE Acoustical Speech and Signal Processing Magazine, Vol. 4, 1987, pp.4–22.
- Ma, W. and Manjunath B., “NETRA: A toolbox for navigating large image database,” In Proc. of the IEEE Int. Conf. on Image Processing, vol. 1, 1997, pp. 568-571.
- Marszałek, M., Laptev, I., Schmid, C., “Actions in Context,” In: CVPR. Florida, 2009.
- Mathias Lux , Jutta Becker , Harald Krottmaier, “ Semantic Annotation and Retrieval of Digital Photos,” In Proc. Of CAiSE Proceedings Information Systems for a Connected Society, 2003.
- Mathias Lux, “Caliph & Emir: MPEG-7 Photo Annotation and Retrieval,” Proceedings of the Seventeen ACM International Conferences on Multimedia, Beijing, China, 2009, pp. 925-926.
- Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt, Sequential deep learning for human action recognition, Proceedings of the Second international conference on Human Behavior Understanding, Springer-Verlag Berlin, Heidelberg, 2011, pp. 29-39.
- Morbidity, Mortality Weekly report (MMWR): Physical Activity Trends— United States, 1990–1998, Vol. 50, 2001.

- N. Chinpanthana, Integrating Qualitative Features with Feature Selection for Semantic Image Classification, International Conference on Management technology and applications (ICMTA2010), Singapore, 10-12 Sept., 2010.
- N. Chinpanthana, Extracting Features with Structural Skeleton Framework for Semantic Image Classification by using Supporting Vector Machine, The 3rd International Conference on Advanced Computer Theory and Engineering, Chengdu, China, 20-22 Aug., 2010.
- Nazli Ikizler-Cinbis, R. Gokberk Cinbis and Stan Sclaroff, Learning Actions From The Web, In International Conference on Computer Vision (ICCV 2009), Kyoto, Japan, September 2009.
- Philippe Mulhem, Joo Hwee Lim, "Symbolic photograph content-based retrieval", *In Proceedings of the eleventh international conference on Information and knowledge management*, McLean, Virginia, USA, 2002, pp. 94 – 101.
- Q. Iqbal and J. K. Aggarwal, CIRES: A System for Content-based Retrieval in Digital Image Libraries, International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, December 2-5, 2002, pp. 205-210.
- Qian Huang, B. Dom, D. Steele, J. Ashley and W. Niblack, "Foreground/Background Segmentation of Color Images by Integration of Multiple Cues", *Proceedings of the IEEE International Conference on Image Processing (ICIP95)*, 1995.
- R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 2nd Ed., Prentice-Hall, Inc., 2002.
- R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Pearson Education Pte. Ltd., 2004.
- R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.
- R. Zhao and W. I. Grosky, "Narrowing the Semantic Gap—Improved Text-Based Web Document Retrieval Using Visual Features", *IEEE Transactions on Multimedia*, Vol. 4, No. 2, 2002, pp. 189-200.
- Rudolph Arnheim, *Art and Visual Perception A Psychology of the Creative Eye*, University of California Press, Ltd., 1974, pp. 11-15.
- Saehoon Yi, Sparse Granger Causality Graphs for Human Action Classification, Pattern Recognition (ICPR), 2012, pp. 3374 – 3377.

Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach, In: International Conference on Pattern Recognition, vol. 3, 2004, pp. 32–36.

Smith, M. *Neural networks for statistical modeling*, New York: Van Nostrand Reinhold, 1993.

T. Kohonen, *Self-organizing maps*, Springer-Verlag, Berlin, 1995.

T. Mitchell, *Machine Learning*, McGraw Hill , 1997.

Tele Tan, Jiayi Chen, Philippe Mulhem and Mohan Kankanhalli, “SmartAlbum – A Multi-Modal Photo Annotation System”, In *Proceedings of the ACM Multimedia*, Juanles-Pins, France, Dec, 2002.

Th. Gevers and A. W. M. Smeulders, “PicToSeek: Combining Colour and Shape Invariant Features for Image Retrieval,” *IEEE Trans. on Image Processing*, vol. 9, no. 1, 2000, pp. 102-119.

The Corbis Corporation: <http://pro.corbis.com> ค้นคืนเมื่อวันที่ 9 ตุลาคม 2555.

Vailaya, A. Figueiredo, M.A.T. Jain, A.K. Hong-Jiang Zhang, “ Image classification for content-based indexing,” *IEEE Transactions on Image Processing*, Vol. 10, issue 1, 2001.

Weilong Yang, Tian Lan, and Greg Mori, SFU at TRECVID 2009: Event Detection, TRECVID Workshop , 2009.

Weilong Yang, Yang Wang, and Greg Mori, Efficient Human Action Detection using a Transferable Distance Function, Ninth Asian Conference on Computer Vision (ACCV), 2009.

Weilong Yang, Yang Wang, and Greg Mori, Human Action Recognition from a Single Clip per Action, 2nd International Workshop on Machine Learning for Vision-based Motion Analysis (at ICCV), 2009.

Weilong Yang, Yang Wang, and Greg Mori, Recognizing Human Actions from Still Images with Latent Poses, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2010.

Xin-Jing Wang, Lei Zhang, Xirong Li, Wei-Ying Ma, “Annotating Images by Mining Image Search Results,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2008.

Y.Wang, H. Jiang, M. S. Drew, Z.-N. Li, and G. Mori. Unsupervised discovery of action classes. In *CVPR*, 2006.

Zhao T, Tang L H, Ip H, Qi F., “Visual Keyword Image Retrieval Based on Synergetic Neural Network for Web-Based Image Search”, *Journal of Real-Time Systems*, Vol.21, Issue 1/2 , 2001, pp.127 – 142.

-----.



ประวัติผู้วิจัย

ผู้ช่วยศาสตราจารย์นศัพธ์ชาณัณ ชินปัญช์ธนะ

สถิติประยุกต์ (คอมพิวเตอร์) สถาบันบัณฑิตพัฒนบริหารศาสตร์
วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยหอการค้าไทย

ปัจจุบัน

อาจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยธุรกิจบัณฑิต

คณะกรรมการกลุ่มผลิตชุดวิชาการเขียนโปรแกรมขั้นสูง

คณะกรรมการกลุ่มผลิตชุดวิชาโครงสร้างข้อมูลและอัลกอริทึม

คณะกรรมการกลุ่มผลิตชุดวิชาการบริหารระบบเครือข่าย

คณะกรรมการกลุ่มผลิตชุดวิชาการระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์

คณะกรรมการกลุ่มผลิตชุดวิชาการจัดการเว็บเซอร์วิส

คณะกรรมการกลุ่มผลิตชุดวิชาการสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์

สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

รางวัลและเกียรติประวัติ

ได้รับทุนการศึกษาประเภทเรียนดี มหาวิทยาลัยหอการค้าไทย พ.ศ. 2534 – 2537

ได้รับปริญญาตรีเกียรตินิยมอันดับ 1 จาก มหาวิทยาลัยหอการค้าไทย

ประสบการณ์งานสอน

อาจารย์ประจำ

มหาวิทยาลัยธุรกิจบัณฑิต

ภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ

- วิชาธุรกิจอิเล็กทรอนิกส์ 2555
- วิชาการพัฒนาและเขียนโปรแกรมบนเว็บ (2550-2554) (Tool: EditPlus, Html, Java script)

- วิชาพื้นฐานสื่อประสมเชิงโต้ตอบ (2553-2554) (Tool: Adobe Photoshop, Illustrator , Flash CS3-5)
- วิชาธุรกิจอิเล็กทรอนิกส์ (1/2553) (Tool: Ms. Office, Dreamwaver)
- วิชาสำนักงานอิเล็กทรอนิกส์ (2550-2552) (Tool: Ms. Office)
- วิชาธุรกรรมอิเล็กทรอนิกส์เบื้องต้น (2/2552) (Tool: Ms. Office , Dreamwaver)
- วิชาการเขียนโปรแกรมเบื้องต้น (2/2550) (Tool: Java program)

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

- วิชาพื้นฐานสื่อประสม (1/2554) (Tool: Adobe Photoshop, Illustrator , Flash CS3-5)
- วิชาการประมวลผลภาพ (Image processing) (2/2552) (Tool: Matlab program)
- วิชาโครงการ (IT project)
 - ระบบวิเคราะห์เปรียบเทียบสเปคเครื่องคอมพิวเตอร์ที่เหมาะสม (Specification and Comparative Analysis for Computer) 2553
 - ระบบวิเคราะห์งบรายได้-รายค่าใช้จ่ายส่วนบุคคล (Analysis Personal Statement of income and Expenditure) 2553
 - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกอาชีพ (Online Career and Education Planning System) 2552
 - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกทัวร์ที่เหมาะสมกับลูกค้า (Intelligent Expert Tour System) 2552
 - โครงการ ระบบวิเคราะห์ดูแลสุขภาพ (Health Care System Analysis) 2551
 - โครงการ ฐานข้อมูลเพื่อเรียกดูเส้นทางในเขตหลักสี่ (Route browsing database for Laksi) 2551

ภาควิชาคอมพิวเตอร์ คณะบริหารธุรกิจ

- วิชาการจัดการฐานข้อมูล (Database management system) (2542-2543)
- วิชาการเขียนโปรแกรมภาษา ซี (1/2542) (Tool: C program)
- วิชาการใช้โปรแกรมสำเร็จรูป (2/2541)
- วิชาคอมพิวเตอร์เบื้องต้น (Introduction to computer) (1/2541)

มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

- วิชาโครงสร้างข้อมูล (Data Structure) (1/2541) (Tool: Borland Turbo C Programming language)

อาจารย์พิเศษและวิทยากรรับเชิญ

- มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี บรรยายและฝึกปฏิบัติ ชุด วิชา การบริหารระบบเครือข่าย
- มหาวิทยาลัยหอการค้าไทย คณะวิศวกรรมศาสตร์ บรรยาย เรื่อง การประยุกต์การใช้งาน Ms. Office PowerPoint และ Visio 2010
- มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี บรรยายและฝึกปฏิบัติ ชุด วิชา การเขียนโปรแกรมเบื้องต้น
- มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี บรรยายและฝึกปฏิบัติ ชุด วิชา ระบบสำนักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์
- มหาวิทยาลัยศรีปทุม บรรยายและปฏิบัติการ วิชาโครงสร้างข้อมูล (Data Structure) (Tool: C++ program)
- มหาวิทยาลัยเกริก บรรยาย วิชาคลังข้อมูล (Data warehouse)
- วิทยาลัยภาคกลาง จ.นครสวรรค์ บรรยาย วิชาความปลอดภัยข้อมูล (Computer Security)

งานวิจัยที่สนใจ

สนใจงานวิจัยทางการค้นคืนความหมายภาพ (Image retrieval) ทั้งภาพธรรมชาติทั่วไป (natural images) และ ภาพที่เป็นส่วนบุคคล (personal images) ในหัวข้อเกี่ยวกับการค้นคืน human activity หรือในหัวข้อการแปลความหมายภาพ (semantic human image) โดยเจาะจงทางด้านการใช้ท่าทางของมนุษย์เพื่อแสดงถึงความหมายของภาพ รวมทั้งการจำแนกภาพต่างๆ เป็นกลุ่ม (Image classification) และการประยุกต์หลักการประมวลผลภาพ เพื่อนำมาใช้ในทางอุตสาหกรรม ได้จริง

หนังสือ

นศัพนธ์ ชินปัญชณะ, หน่วยที่ 1 ความรู้เบื้องต้นเกี่ยวกับเครือข่ายคอมพิวเตอร์ และ หน่วยที่ 3 ชุมชนเครือข่าย, หลักการบริหารและจัดการเครือข่าย (Network Management) (99412) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2556.

นศัพนธ์ ชินปัญชณะ, หน่วยที่ 12 โลจิสติกส์และโซ่อุปทานในระบบพาณิชย์อิเล็กทรอนิกส์ และ หน่วยที่ 13 การชำระเงินในระบบพาณิชย์อิเล็กทรอนิกส์, ระบบสำนักงานอัตโนมัติและ

พาณิชย์อิเล็กทรอนิกส์ (Office Automation System and Electronic Commerce)
(99311) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2555.

นิต์พชาณัณ ชินปัญชัณนะ, หน่วยที่ 3 แบบจำลองเครือข่ายและโพรโทคอลกับการบริการผ่าน
เว็บ และ หน่วยที่ 5 ความรู้พื้นฐานของภาษาเอ็กซ์เอ็มแอล, เทคโนโลยีการบริการผ่านเว็บและ
การประยุกต์ (Web Services Technology and Applications) (99301)
มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2555.

นิต์พชาณัณ ชินปัญชัณนะ, หน่วยที่ 3 ตัวกลางในการสื่อสารข้อมูล และ หน่วยที่ 10 โพรโท
คอล,การสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์ (Data Communications and
Networking), มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2553.

เอกสารประกอบการสอน

นิต์พชาณัณ ชินปัญชัณนะ, เอกสารประกอบการสอนวิชา การประมวลผลภาพ (Image
processing), มหาวิทยาลัยธุรกิจบัณฑิตย์, 2552.

งานวิจัย

นิต์พชาณัณ ชินปัญชัณนะ, การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบ
จับคู่, มหาวิทยาลัยธุรกิจบัณฑิตย์, 2554.

นิต์พชาณัณ ชินปัญชัณนะ, ระบบตรวจนับวัตถุแบบอัตโนมัติด้วยเทมเพลตแมชชีง,
มหาวิทยาลัยธุรกิจบัณฑิตย์, 2553.

นิต์พชาณัณ ชินปัญชัณนะ, การจำแนกความหมายของภาพจากวัตถุโดยใช้หลักการโครงสร้าง
สเกตริตรอน, มหาวิทยาลัยธุรกิจบัณฑิตย์, 2552.

ผลงานทางวิชาการ

N. Chinpanthana, Semantic Similarity Measure with Conceptual Graph-Based
Image Annotation, International Conference on Advanced Computer Science
Application and Technologies (ASCAT 2012), Kuala Lumpur, Malaysia, Nov., 2012.

N. Chinpanthana and T. Phiasai., Automatic Counting System With Normalized Correlation Coefficient Template Matching, International Conference on Computer and Information Technology, Amsterdam, Netherlands, July 13-15, 2011.

N. Chinpanthana, Integrating Qualitative Features with Feature Selection for Semantic Image Classification, International Conference on Management technology and applications (ICMTA2010), Singapore, 10-12 Sept., 2010.

N. Chinpanthana and T. Phiasai., Multi-Layer Perception Networks for Semantic Image Classification with Structural Skeleton Framework, International Technical Conference on Circuit/Systems Computers and Communications, Pattaya, Thailand, July., 2010.

N. Chinpanthana, Extracting Features with Structural Skeleton Framework for Semantic Image Classification by using Supporting Vector Machine, The 3rd International Conference on Advanced Computer Theory and Engineering, Chengdu, China, 20-22 Aug., 2010.

N. Chinpanthana, Semantic Salient Images Based on Similarity Matching with Conceptual Graph, International Technical Conference on Circuits/Systems, Computers and Communications, Jeju, Korea 2009.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "Semantic Human Image Classification Based on Energy Action Model with Essential Reference points," Int. Symposium on Communications and Information Technologies, 16-19 Oct, Sydney, AUS, 2007.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "High-Level Semantic Image Classification by Using Energy Expenditure," 2007 International Workshop on Smart Info-Media Systems in Bangkok, November 1-2, 2007, Bangkok, Thailand.

S. Chinpanchana, "Semantic Human Action Classification Based on Energy-Action Model", Tencon 2006 IEEE Region 10, Hongkong, China, 2006.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "Semantic Personal Image Classification by Energy Expenditure," Int. Symposium on Communications and Information Technologies, Beijing, China, 2005.

S. Chinpanchana, S. Maneewongvatana and B. Thipakorn, "Semantic Personal Image Pattern Classification Based on Human Body," Asia Information Retrieval Symposium, Beijing, China, Oct. 2004.

S. Chinpanchana and B. Thipakorn, "Semantic Classification of Personal Images Based on Human Action and Associate Bayesian Rule," International Technical Conference on Circuits/Systems, Computers and Communications, Sendai, Japan, 2004.

