



รายงานผลการวิจัย

เรื่อง

การจำแนกความหมายภาพด้วยการใช้ความสัมพันธ์แอ็คชันของวัตถุในภาพ สำหรับช่วยเหลือ
ผู้สูงอายุ

SEMANTIC IMAGE CLASSIFICATION WITH RELATIONSHIP BETWEEN OBJECTS FOR
SAFETY OLDER ADULTS

โดย

นศภัชชาณัน ชินปัญชธนะ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

2562

ชื่อเรื่อง : การจำแนกความหมายภาพด้วยการใช้ความสัมพันธ์แอ็คชันของวัตถุในภาพ สำหรับช่วยเหลือผู้สูงวัย

ผู้วิจัย : นศภัชชาณัน ชินปัญญาชนะ สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิต

ปีที่พิมพ์ : 2563 สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิต

แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิต จำนวนหน้างานวิจัย 52 หน้า

คำสำคัญ : การประมวลผลภาพ, การจำแนกความหมายภาพ, ผู้สูงวัย ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิต

บทคัดย่อ

ปัจจุบันการดูแลผู้สูงอายุเป็นสิ่งที่จำเป็นอย่างมาก การตรวจจับกิจกรรมประจำวันกลายเป็นหัวข้อวิจัยที่น่าสนใจ จากการทบทวนวรรณกรรมได้มีการแสดง ถึงการรวบกันของรูปแบบทางสถิติเพื่อทำการแมพไปยังความสัมพันธ์ของภาพ มีหลายวิธีการที่ถูกนำเสนอเพื่อใช้ในการระบุกิจกรรมด้วยคุณลักษณะต่างๆบางแอปพลิเคชันที่จำเป็นต้องทำความเข้าใจข้อมูลภาพก่อนด้วยการจำแนกข้อมูลวัตถุบนภาพ แต่อย่างไรก็ตามยังคงเป็นหัวข้อที่ท้าทายเพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากขึ้น ดังนั้นในงานวิจัยนี้ได้นำเสนอการจำแนกข้อมูลผู้สูงวัยที่ขึ้นกับวัตถุที่อยู่ในสภาพแวดล้อมและความสัมพันธ์ โดยที่มี 3 ความสัมพันธ์ ประกอบด้วย implied-by, type-of และ mutually exclusive มีการแบ่งลำดับขั้นตอนวิธีการเป็น 3 ส่วนหลักดังนี้ (1) การเตรียมข้อมูล (2) การทำนายความสัมพันธ์ (3) การวัดและการประเมินผลการทำงาน โดยทดสอบด้วยข้อมูลจากฐานข้อมูลภาพและมีการแสดงประสิทธิภาพการทำงานด้วยค่าเฉลี่ยความแม่นยำมากถึง 75.69%

research topic : SEMANTIC IMAGE CLASSIFICATION WITH RELATIONSHIP BETWEEN OBJECTS
FOR SAFETY OLDER ADULTS

Researcher: Nutchanut Chinpanthana Institution: Dhurakijbundit University

Year of Publication : 2020

Place of publication: Dhurakijbundit University

place of collection: Dhurakijpundit University Research Center Number of page :52 pages

Keyword: Image Processing, Classification, Older adults Copyright : Dhurakijbundit University

Abstract

As elderly care is getting more and more important, monitoring of activity of daily living has become an active research topic. We have previously shown that the combination of a visual model and a statistical semantic prior model can improve on the task of mapping images to their associated scene description. Multiple techniques were proposed to identify activities, each with their features, advantages and limits. Applications require an understanding of an activity of daily living that goes beyond the simple detection and classification of its objects. However, it is a very challenging issue and none of the existing methods provides robust results. In this paper, we propose to classify an elderly activity recognition process by relying on environment objects and relationships. The interaction relationships are including implied-by, type-of and mutually exclusive. The approach is composed of four main phases: (1) data preprocessing (2) predict relations of actions (3) Measurement and Evaluation. We train and test our model on a large scale image dataset of relationship actions. The experimental results indicate that our proposed approach offers significant performance improvements in the classification of relationship actions with maximum success rate of 75.69%

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จลุล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และแรงผลักดัน และอีกหลายฝ่ายที่ได้ให้ความช่วยเหลือเอื้อเฟื้อที่เป็นประโยชน์ในการทดลองโปรแกรม และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ผู้ทรงคุณวุฒิ ศ.ดร.โกสินทร์ จ่านงไท, ผศ.ดร.ณรงค์เดช กิริติพรานนท์, ดร.เตชรัฐศิณป์ เพี้ยซ้าย ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิต ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วย

นศัพชาณันณ ชินปัญชธนะ

กันยายน 2562



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	(1)
สารบัญตาราง	(10)
สารบัญรูปภาพ	(12)
บทที่ 1	
บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์ของการวิจัย	2
สมมติฐานการวิจัย	2
ขอบเขตของการวิจัย	2
ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2	
งานวิจัยที่เกี่ยวข้อง	4
2.1 การค้นคืนข้อมูลภาพด้วยพีเจอร์ระดับต่ำ	4
2.2 การค้นคืนด้วยการให้ความหมายภาพ	6
2.3 การค้นคืนด้วยกิจกรรมมนุษย์	9
บทที่ 3	
การดำเนินงานวิจัย	14
3.1 การเตรียมข้อมูล	14
3.2 การทำนายความสัมพันธ์แอ็คชัน	20
3.3 การวัดและการประเมินผลการทำงาน	23
บทที่ 4	
ผลการวิจัย	30
4.1 การวัดประสิทธิภาพการจำแนกข้อมูลภาพจากฐานข้อมูลมาตรฐาน Ground Truth	30
4.2 การวัดประสิทธิภาพการจำแนกกิจกรรมด้วยความสัมพันธ์ระหว่างวัตถุและแอ็คชัน	31
4.3 การวัดประสิทธิภาพจากการนำไปประยุกต์ใช้กับการเฝ้าระวังและช่วยเหลือผู้สูงอายุ	35
บทที่ 5	
สรุป อภิปราย และข้อเสนอแนะ	37
สรุปผลการวิจัย	37
อภิปรายผล	38
ข้อเสนอแนะ	39
บรรณานุกรม	40
ภาคผนวก	
ก. ประวัติผู้วิจัย	

สารบัญตาราง

ตารางที่	หน้า
4.1 ผลการจำแนกข้อมูลภาพกับวิธีมาตรฐานด้วยข้อมูล Ground Truth	31
4.2 ผลการจำแนกกิจกรรมด้วยความสัมพันธ์ของวัตถุและแอ็คชันด้วยชุดทดลอง Set I	33
4.3 ผลการจำแนกกิจกรรมด้วยความสัมพันธ์ของวัตถุและแอ็คชัน ชุดทดลอง Set II	34
4.4 ผลการจำแนกข้อมูลภาพด้วยข้อมูลผู้สูงวัย ชุดทดลอง Set I	35
4.5 ผลการจำแนกข้อมูลภาพด้วยข้อมูลผู้สูงวัย ชุดทดลอง Set II	35



สารบัญรูปภาพ

ภาพที่	หน้า
2.1 ภาพตัวอย่างของการค้นคืนที่ถูกแปลงเป็นพีเจอร์ระดับต่ำ	4
2.2 แสดงตัวอย่างกลุ่มภาพผลลัพธ์จากการประมวลผลภาพระดับต่ำ	5
2.3 แสดงตัวอย่างกลุ่มภาพผลลัพธ์จากการประมวลผลภาพระดับต่ำ	6
2.4 ภาพตัวอย่างต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นพีเจอร์ระดับต่ำ	6
2.5 ตัวอย่างการแทนความหมายของวัตถุและความสัมพันธ์ด้วยคำศัพท์	7
2.6 แสดงการใช้ความสัมพันธ์ของกราฟเพื่อหาความหมายภาพ	7
2.7 ตัวอย่างการแทนของความสัมพันธ์ระหว่างวัตถุและรูปแบบความสัมพันธ์	8
2.8 ตัวอย่างการนำคำศัพท์มาสร้างความสัมพันธ์ที่ถูกแท็กในฐานข้อมูลภาพ	8
2.9 ตัวอย่างการจำแนกท่าทางมนุษย์	10
2.10 แสดงตัวอย่างกระบวนการรู้จำกิจกรรมของมนุษย์	12
3.1 ขั้นตอนการจำแนกความหมายภาพด้วยการทำนายความสัมพันธ์คำศัพท์บนภาพ	14
3.2 ตัวอย่างการแบ่งส่วนในฐานข้อมูล Pascal2	15
3.3 ตัวอย่างการจำแนกแอ็คชันบนฐานข้อมูล Pascal2	16
3.4 ตัวอย่างจำลอง Synset ในอนุกรมวิธาน WordNet	17
3.5 รายการคำหลักเริ่มต้นบน WordNet	18
3.6 ตัวอย่างคำศัพท์ที่ประกอบด้วยคำนามและแอ็คชันบนฐานข้อมูลภาพ	19
3.7 ความสัมพันธ์ระหว่างวัตถุและท่าทางภายในภาพ	19
3.8 ขั้นตอนการสร้างความสัมพันธ์ของแอ็คชัน	21
3.9 ตัวอย่างเซตความสัมพันธ์ของการแอ็คชันของข้อมูลภาพ	23
3.10 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง	24
3.11 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน	25
3.12 Algorithm Stochastic Sub-Gradient Projection	28
4.1 ตัวอย่างภาพในกลุ่มกิจกรรม	32
5.1 ตัวอย่างภาพผู้สูงอายุที่ได้รับการดูแล และควรเฝ้าระวัง	38

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การประมวลผลปัจจุบันความก้าวหน้าของอุปกรณ์อิเล็กทรอนิกส์รวมทั้งเทคโนโลยีได้พัฒนาไปอย่างรวดเร็ว และราคาอุปกรณ์ถูกลง ขนาดหน่วยความจำเพิ่มมากขึ้น การจัดเก็บข้อมูลมีขนาดใหญ่ขึ้นตามจำนวนข้อมูลที่เพิ่มขึ้น เช่นเดียวกันกับการพัฒนาการถ่ายภาพที่มีจำนวนเพิ่มมากขึ้นอย่างรวดเร็วเช่นเดียวกัน ทำให้การจำแนกข้อมูลภาพบนฐานข้อมูลขนาดใหญ่ยังคงเป็นหัวข้องานวิจัยทางด้านประมวลผลภาพ (Image Processing) ที่มีความท้าทายอย่างมาก ทุกอย่างบนฐานข้อมูลส่วนใหญ่จะมีการเก็บในรูปแบบทั้งข้อความและข้อมูลภาพ และสิ่งที่ยากคือการจำแนกข้อมูลภาพอย่างไรเพื่อให้ได้ข้อมูลภาพตรงตามความหมายของผู้ใช้งาน การอย่างแท้จริง มีนักวิจัยพยายามใช้วิธีการเรียนรู้จากข้อมูลตัวอย่าง ทั้งในรูปแบบของการนำข้อมูลระดับต่ำมาใช้ในการค้นคืนภาพที่เรียกว่า Content-Based Image Retrieval [1] และได้มีการปรับปรุงพัฒนาในรูปแบบของการใช้คำศัพท์เพื่อเป็นข้อมูลสำหรับการค้นคืนแทน และถูกเรียกว่า Annotation-Based Image Retrieval [2] แต่อย่างไรก็ตามยังคงมีการพัฒนาโดยนำวิธีการแทนข้อมูลด้วยกราฟแบบลำดับชั้น [3, 4] หรือการวัดความคล้ายด้วยความหมายของคำศัพท์ รวมทั้งวิธีการอื่นที่ซับซ้อนเพื่อนำมาประยุกต์ใช้แต่ยังไม่สามารถได้ภาพผลลัพธ์ตามความต้องการ

งานวิจัยสำหรับการค้นคืนความหมายภาพเพื่อให้ได้ผลลัพธ์ตามความต้องการของผู้ใช้งานสามารถแบ่งรูปแบบวิธีการค้นคืนความหมายภาพได้ดังนี้ (1) การค้นคืนข้อมูลภาพด้วยข้อมูลระดับต่ำ (Content-Based Image Retrieval) [1, 3] การแทนค่าข้อมูลวัตถุภายในภาพด้วยข้อมูลระดับต่ำ (Low-Level Features) [5] ที่ประกอบด้วย สี, รูปร่าง หรือ ลวดลาย ที่นำมารวมกันเพื่อใช้เป็นคุณสมบัติสำหรับการค้นคืนภาพ แต่ผลลัพธ์ที่ได้จากการค้นคืนด้วยข้อมูลภาพระดับต่ำจะมีความเหมือนกันทางกายภาพเท่านั้น จะไม่ได้ความหมายภาพที่ต้องการและได้มีการปรับปรุงพัฒนาในรูปแบบของการใช้คำศัพท์เพื่อเป็นข้อมูลสำหรับการค้นคืนแทน ถูกเรียกว่า (2) การค้นคืนข้อมูลภาพด้วยข้อมูลคำศัพท์ (Annotation-Based Image Retrieval) [6] การแทนค่าข้อมูลวัตถุภายในภาพด้วยคำศัพท์ (keyword) เช่น “sky”, “bicycle”, “home” เป็นต้น โดยทำการให้ความหมายภาพเป็นวัตถุที่เด่นชัดบนภาพดังแสดงในรูปที่ 1 เพื่อใช้สำหรับการค้นคืนข้อมูลภาพแทนข้อมูลระดับต่ำ ผลลัพธ์ที่ได้จะมีความเหมือนกันกับข้อมูลคำศัพท์ที่ใช้ในการค้นหา แต่อย่างไรก็ตาม งานวิจัยส่วนใหญ่พยายามปรับปรุงรูปแบบของผลลัพธ์ของการจำแนกที่ให้ความหมายตรงตามความต้องการของผู้ใช้งานที่เรียกว่า ความหมายภาพ (Semantic Image) ด้วยการสร้างความสัมพันธ์ของวัตถุภายในภาพ (Spatial Relationships) เช่น “below”, “above”, “hold” เป็นต้น [7] เพื่อหากิจกรรมที่เกิดขึ้นภายในภาพ โดยอาศัยการเรียนรู้เพื่อให้เกิดความสัมพันธ์ของ “person” และ “object” ภายใน แต่อย่างไรก็ตามบางกลุ่มได้พยายามสร้างวิธีการด้วยโครงสร้างความหมายแบบลำดับในเชิงการมองเห็น (Semantic Hierarchy for Vision) [8, 9] กับโครงสร้างของการเรียนรู้รูปแบบเพื่อจำแนกกลุ่มและแบ่งส่วนภาพ เพื่อสร้างความสัมพันธ์ในรูปแบบที่กำหนด และหาความสัมพันธ์ของภาพจากคำศัพท์ภายใน ด้วยการเรียนรู้ทำให้ครอบคลุมรูปแบบการจำแนก

ภาพ Deng et al. [10], ใช้กราฟเพื่อสร้างความสัมพันธ์ (DAG relationships) และ Mutual exclusion ระหว่างข้อมูลสำหรับการจำแนกที่ดีขึ้น บางกลุ่มพยายามที่จะมีการใช้กลุ่มข้อมูลที่แตกต่างกันเพื่อทำการเรียนรู้ รูปแบบความสัมพันธ์ที่เป็นแบบ ลำดับชั้นบนพื้นฐานของการวัดความคล้ายและการเกิดขึ้นของความสัมพันธ์ [11] แต่เมื่อมีการเรียนรู้เกิดขึ้นทำให้มีการสร้างองค์ความรู้เพื่อเก็บเป็นข้อมูลคำศัพท์บางที่มาจากเว็บไซต์ที่สร้างขึ้นเพื่อรวบรวมคำศัพท์ที่เกี่ยวข้อง แต่อย่างไรก็ตามรูปแบบของการเรียนรู้แบบอัตโนมัติยังคงมีความจำเป็น ดังนั้นในงานวิจัยนี้จึงได้นำเสนอวิธีการจำแนกข้อมูลแบบด้วยการสร้างพารามิเตอร์ของความสัมพันธ์ระหว่างวัตถุและแอ็คชันจากข้อมูลภายในภาพเพื่อสามารถนำไปใช้ในการหาภาพผลลัพธ์ที่ต้องการของผู้ใช้งานได้มากขึ้น

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแบบจำลองวิเคราะห์การค้นหาคำความหมายภาพรูปแบบใหม่
2. เพื่อใช้เทคนิคใหม่ที่น่าสนใจมาใช้ในการค้นหาคำความเหมือนกันของภาพดิจิทัล
3. เพื่อพัฒนาเทคนิคในการหาความเหมือนกันของภาพดิจิทัล
4. เพื่อแจ้งเฝ้าระวังสำหรับผู้สูงอายุที่ต้องดูแลเป็นพิเศษ

สมมติฐานการวิจัย

การใช้วิธีการแทนข้อมูลด้วยความสัมพันธ์ของแอ็คชันที่เกิดขึ้นระหว่างวัตถุและแอ็คชัน การทำนายด้วยความสัมพันธ์ของข้อมูลภายในที่ประกอบด้วย implied-by, type-of และ mutually exclusive วิธีการนี้สามารถนำมาใช้ในการจำแนกข้อมูลภาพให้ลงในกลุ่มความหมายได้มากกว่าวิธีการอื่น

ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัลที่เห็นวัตถุชัดเจน
2. ข้อมูลคำหลักบนภาพจะถูกจัดเก็บเป็นข้อมูลเข้าก่อนเริ่มการประมวลผล
3. ความหมายของภาพต้องชัดเจน

นิยามตัวแปร

1. การประมวลผลภาพ (image processing) หมายถึง การนำภาพมาผ่านกระบวนการ เพื่อประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง (ภาพถ่าย) หรือภาพวิดีโอ (วีดีโอ) และนำมาใช้งาน
2. การค้นคืนภาพ (image retrieval) หมายถึง การสืบค้นภาพจากฐานข้อมูลภาพดิจิทัลเพื่อให้ได้ภาพที่ต้องการ
3. คุณลักษณะข้อมูล (feature) หมายถึง เป็นพีเจอร์ หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป

- 4.วัตถุ (object) หมายถึง ส่วนของวัตถุนภาพ เช่น คน, ต้นไม้, รถยนต์ เป็นต้น
- 5.คำหลัก (keyword) หมายถึง คำที่มีความหมายได้ใจความใช้แทนวัตถุนภาพ
- 6.แอ็คชั่น (action) หมายถึง การกระทำ

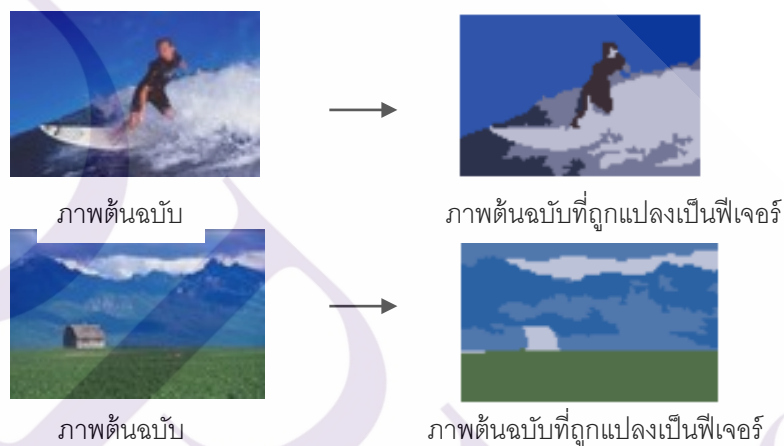
ประโยชน์ที่คาดว่าจะได้รับ

- 1.เพื่อให้ได้แนวทางการแปลและการตีความหมายภาพ
- 2.ช่วยให้การใช้คำศัพท์ในการค้นหาข้อมูลภาพได้ผลตามความหมายของภาพมากขึ้น
- 3.สามารถนำส่วน Algorithm มาประยุกต์ใช้เพื่อเฝ้าระวังสำหรับผู้สูงวัยได้
- 4.สามารถนำส่วน Algorithm เข้ามาประยุกต์ใช้ในการค้นหาข้อมูลภาพตามเป้าหมายที่ตั้งไว้ในระบบการค้นหาข้อมูลได้ ไม่ว่าจะเป็นทาง ธุรกิจ, อุตสาหกรรม เป็นต้น



บทที่ 2 งานวิจัยที่เกี่ยวข้อง

ปัจจุบันความก้าวหน้าของอุปกรณ์อิเล็กทรอนิกส์รวมทั้งเทคโนโลยีได้พัฒนาไปอย่างรวดเร็ว และราคาอุปกรณ์ถูกลง ขนาดหน่วยความจำเพิ่มมากขึ้น การจัดเก็บข้อมูลมีขนาดใหญ่ขึ้นตามจำนวนข้อมูลที่เพิ่มขึ้น เช่นเดียวกันกับการพัฒนาการของภาพถ่ายที่มีจำนวนเพิ่มมากขึ้นอย่างรวดเร็วเช่นเดียวกัน ทำให้การจำแนกข้อมูลภาพบนฐานข้อมูลขนาดใหญ่ยังคงเป็นหัวข้องานวิจัยทางการประมวลผลภาพ (Image Processing) ที่มีความท้าทายอย่างมาก ทุกอย่างบนฐานข้อมูลส่วนใหญ่จะมีการเก็บในรูปแบบทั้งข้อความและข้อมูลภาพ และสิ่งที่ยากคือการจำแนกข้อมูลภาพอย่างไรเพื่อให้ได้ข้อมูลภาพตรงตามความหมายของผู้ใช้งานการอย่างแท้จริง



ภาพที่ 2.1 ภาพตัวอย่างของการค้นคืนที่ถูกแปลงเป็นพีเจอร์ระดับต่ำ

ที่มา <http://simplr.com/cgi-bin/zwang/>

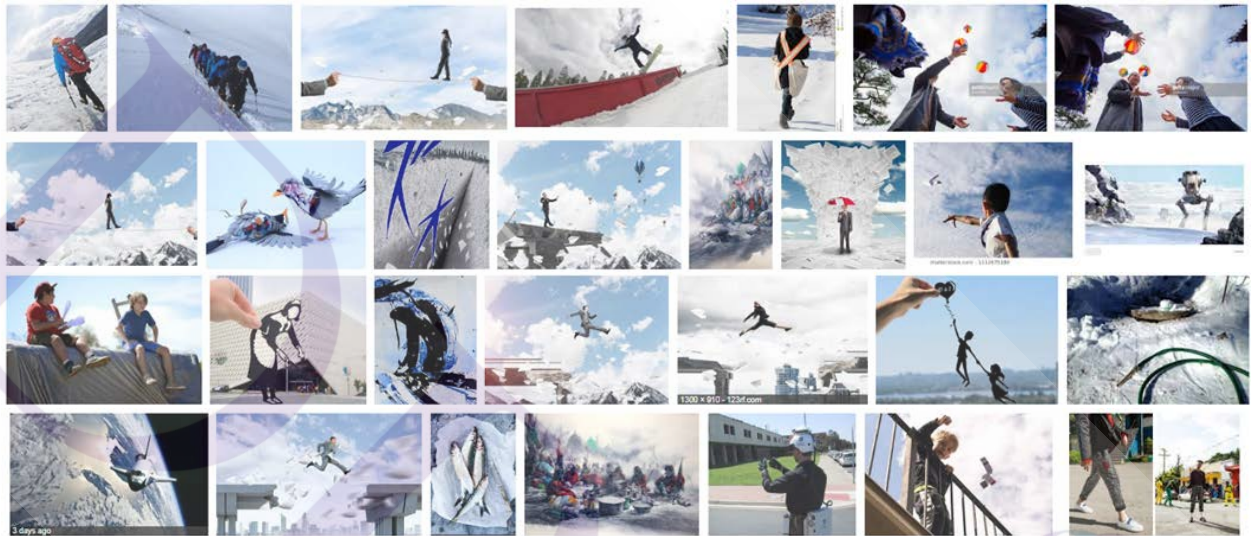
2.1 การค้นคืนข้อมูลภาพด้วยพีเจอร์ระดับต่ำ

การเลือกพีเจอร์หรือคุณลักษณะของข้อมูลจากภาพเป็นสิ่งจำเป็นเพื่อช่วยทำให้การคาดคะเนเพื่อให้ได้ข้อมูลภาพที่ต้องการต้องเป็นการค้นคืนความหมายภาพเพื่อให้ได้ผลลัพธ์ตามความต้องการของผู้ใช้งาน จากวิธีการในอดีตไปนี่จะเป็นการแนะนำรูปแบบของพีเจอร์ต่างๆ ของพีเจอร์ระดับต่ำ เป็นการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ เช่น สี (color) ลวดลาย (texture) รูปทรง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) เป็นกระบวนการที่สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานเมื่อนำมาจำแนกด้วยพีเจอร์ระดับต่ำ (low-level feature) ภาพผลลัพธ์ส่วนใหญ่เป็นภาพที่มีคุณลักษณะไม่ซับซ้อน มีรูปทรงที่แตกต่างกันอย่างเด่นชัด ด้วยคุณลักษณะพีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากภาพที่ 2.1 เป็นงานวิจัยในยุคแรกที่มีการนำการประมวลผลด้วยพีเจอร์ระดับต่ำที่เป็นสี เป็นหลัก ชื่อว่า Semantics-sensitive Integrated Matching for Picture Libraries (Simplicity) ผลลัพธ์ของกลุ่มภาพที่ค้นหาโดยรวมแล้วจะมีความคล้ายคลึงของสีเสียส่วนใหญ่

ใหญ่ ดังนั้นผลลัพธ์ของการค้นคืนภาพส่วนใหญ่จะมีโทนที่ในรูปแบบเดียวกัน ดังแสดงในภาพที่ 2.2 เมื่อต้องการค้นคืนภาพที่มีลักษณะดังแสดงในภาพ จะเห็นว่าผลลัพธ์ของการค้นคืนภาพจะมีโทนสีใกล้เคียงกันแต่ความหมายของภาพหรือวัตถุที่แสดงผลจะมีความต่างกัน



ภาพตั้งต้นเพื่อทำการค้นคืน



ภาพผลลัพธ์จากการค้นคืนด้วยคุณลักษณะต่ำ

ภาพที่ 2.2 แสดงตัวอย่างกลุ่มภาพผลลัพธ์จากการประมวลผลภาพระดับต่ำ[M. Everingham 2010]

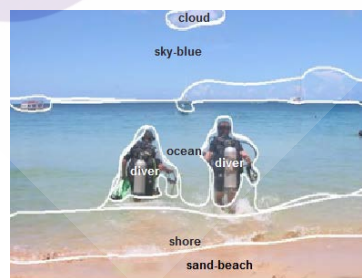
การค้นคืนภาพด้วยการใช้ฟีเจอร์ระดับต่ำหลายแบบผสมกัน ผลลัพธ์ที่เกิดจากงานวิจัยประเภทนี้จะได้ภาพของความคล้ายคลึงจากลักษณะทางกายภาพโดยรวม ดังแสดงในภาพที่ 2.3 จากการค้นคืนภาพด้วยภาพเริ่มต้นด้านมุมซ้ายสุด คือ ภาพที่ประกอบด้วยคน และ ม้า ที่แสดงความสัมพันธ์กันในลักษณะสัมพันธ์ตัวม้า แต่จากกลุ่มภาพผลลัพธ์จะสังเกตว่าเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก ภาพลำดับแรกจะมีวัตถุที่เหมือนกันคือคนและ ม้า แต่มีลักษณะวัตถุที่แตกต่างกันนอกนั้นจะมีคนและม้าบ้าง ในลำดับภาพที่ 7 มีการสื่อความหมายในรูปแบบคล้ายกันของวัตถุ การประมวลผลภาพด้วยข้อมูลภาพจากการสกัดข้อมูลภาพระดับต่ำมาใช้อย่างคงมีการปรับปรุงเทคนิคอยู่เสมอไม่ว่าจะเป็นการผสมผสานฟีเจอร์ต่างๆเข้าด้วยกัน ปรับปรุงอัลกอริทึมเพื่อให้เกิดผลลัพธ์ที่ดีขึ้น การรวมเทคนิคด้วยคุณลักษณะสีและรูปร่างของภาพเพื่อทำการค้นคืนภาพ



ภาพที่ 2.3 แสดงตัวอย่างกลุ่มภาพผลลัพธ์จากการประมวลผลภาพระดับต่ำ ภาพมุมซ้ายบน เป็นภาพเริ่มต้นค้นคืน

2.2 การค้นคืนด้วยการให้ความหมายภาพ

ภาพทุกภาพมีความหมายภายในตัวเอง ปัจจุบันภาพมีจำนวนค่อนข้างเยอะการเก็บภาพส่วนใหญ่จะถูกจัดกลุ่มตามความต้องการของผู้เก็บและเรียกใช้งาน แต่โดยทั่วไปการเก็บข้อมูลภาพมักจะใช้ข้อมูลจากใบหน้าคนบนภาพถ่ายหรือสิ่งที่อยู่นภาพ เช่น สัตว์, สิ่งของ, วัตถุเด่น เป็นต้น ดังนั้นมีงานวิจัยที่ใช้การรู้จำใบหน้าคนเพื่อระบุตัวตนของบุคคลในภาพ หรือ การบอกรหัสที่อยู่นภาพ เช่น สัตว์ สิ่งของ ที่มีขนาดเด่นชัด เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นด้วยกลุ่มคำศัพท์เป็น การประมวลผลภาพระดับสูง (high level image processing) เป็นการใช้คำศัพท์เพื่อเป็นข้อมูลสำหรับการค้นคืนแทน เรียกว่า Annotation-Based Image Retrieval งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุที่มีความหมาย และแทนวัตถุนั้นๆ ด้วยคำศัพท์บนภาพ หรือการให้ความหมายของวัตถุนบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass” , “plant” , “boat”, “sky” เป็นต้น [Escalante, H. J 2009] ดังแสดงในภาพที่ 2.4



ภาพที่ 2.4 ภาพตัวอย่างต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นพีเจอร์ระดับต่ำ
ที่มา <https://www.imageclef.org/SIAPRdata>



Ground truth sky, grass, horse, woman
Top 3 key sky, horse, woman
Relationship (horse, hold, woman)



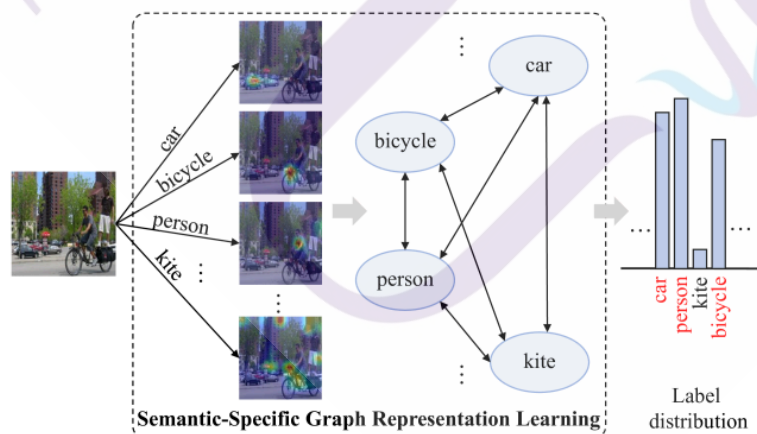
Ground truth kid, dog
Top 3 key kid, dog, building
Relationship (kid, kiss, dog)



Ground truth woman
Top 3 key woman, stadium, track
Relationship (woman, run, track)

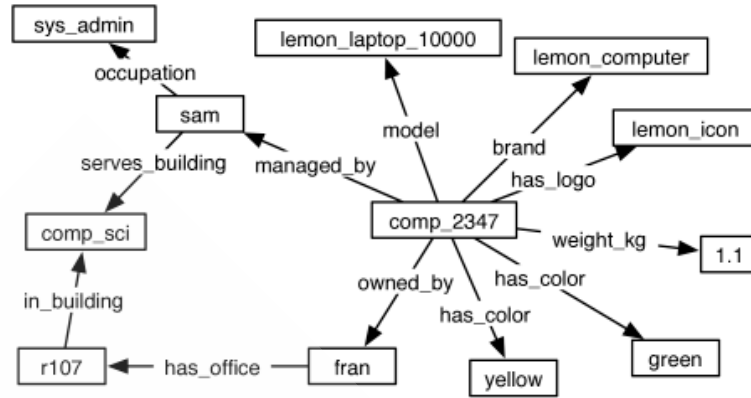
ภาพที่ 2.5 ตัวอย่างการแทนความหมายของวัตถุและความสัมพันธ์ด้วยคำศัพท์

คำศัพท์บนภาพถูกจัดเก็บลงบนฐานข้อมูลดังแสดงในภาพที่ 2.5 และมีการใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมายของคำศัพท์ที่มีความสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำศัพท์ เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น จะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด ได้สร้างความสัมพันธ์ของกลุ่มคำศัพท์โดยจัดหมวดหมู่ของคำศัพท์ที่มีความเกี่ยวข้องเข้าด้วยกัน และได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใดและได้มีการปรับปรุงพัฒนาในรูปแบบอย่างต่อเนื่องแต่อย่างไรก็ตามยังคงมีการพัฒนาโดยนำวิธีการแทนข้อมูลภาพด้วยความสัมพันธ์ของกราฟเพื่อให้เกิดความหมายของภาพ รวมทั้งวิธีการอื่นที่ซับซ้อนเพื่อนำมาประยุกต์ใช้แต่ยังไม่สามารถได้ภาพผลลัพธ์ตามความต้องการ [Tianshui Che 2019] ดังแสดงในภาพที่ 2.6



ภาพที่ 2.6 แสดงการใช้ความสัมพันธ์ของกราฟเพื่อหาความหมายภาพ

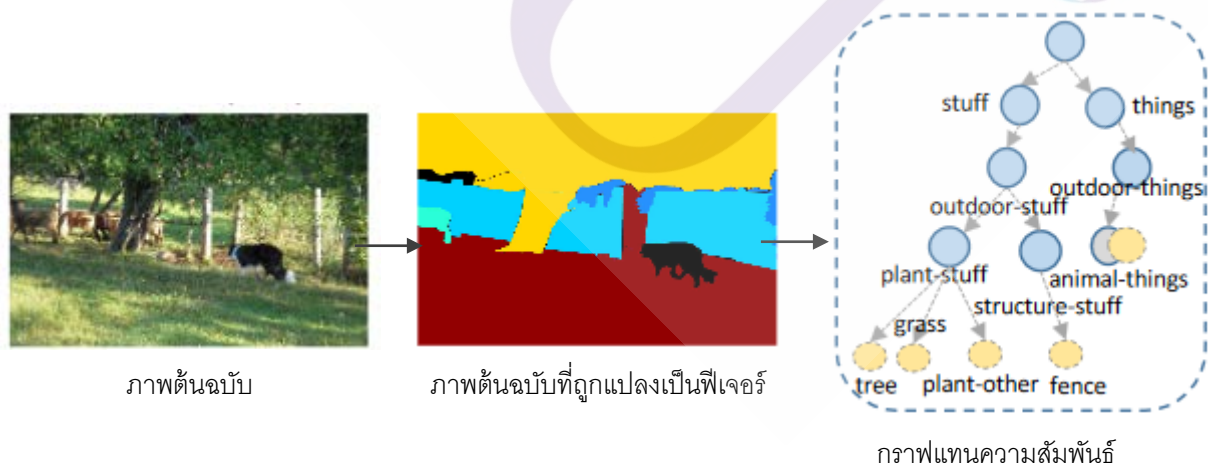
<https://www.groundai.com/project/learning-semantic-specific-graph-representation-for-multi-label-image-recognition/>



ภาพที่ 2.7 ตัวอย่างการแทนของความสัมพันธ์ระหว่างวัตถุและรูปแบบความสัมพันธ์

<https://artint.info/2e/html/ArtInt2e.Ch14.S2.SS2.html>

การสร้างความสัมพันธ์ของวัตถุภายในภาพ เช่น “below”, “above”, “hold” เป็นต้น เพื่อหากิจกรรมที่เกิดขึ้นภายในภาพ ดังแสดงในภาพที่ 2.7 โดยอาศัยการเรียนรู้เพื่อให้เกิดความสัมพันธ์ของ “person” และ “object” ภายใน แต่อย่างไรก็ตามบางกลุ่มได้พยายามสร้างวิธีการด้วยโครงสร้างความหมายแบบลำดับในเชิงการมองเห็น กับโครงสร้างของการเรียนรู้รูปแบบเพื่อจำแนกกลุ่มและแบ่งส่วนภาพ เพื่อสร้างความสัมพันธ์ในรูปแบบที่กำหนด และหาความสัมพันธ์ของภาพจากคำศัพท์ภายใน ด้วยการเรียนรู้ทำให้ครอบคลุมรูปแบบการจำแนกภาพ ใช้กราฟเพื่อสร้างความสัมพันธ์ ระหว่างข้อมูลสำหรับการจำแนกที่ดีขึ้น บางกลุ่มพยายามที่จะมีการใช้กลุ่มข้อมูลที่แตกต่างกันเพื่อทำการเรียนรู้ รูปแบบความสัมพันธ์ที่เป็นแบบลำดับขั้นบนพื้นฐานของการวัดความคล้ายและการเกิดขึ้นของความสัมพันธ์ [Xiaodan Liang 2018] ดังแสดงในภาพที่ 2.8 ตัวอย่างการนำคำศัพท์มาสร้างความสัมพันธ์ที่ถูกแก้ในฐานข้อมูลภาพ โดยโครงสร้างที่เกิดขึ้นเป็นความสัมพันธ์ของความหมายของวัตถุในรูปแบบของออนโทโลยี แต่เมื่อมีการเรียนรู้เกิดขึ้นทำให้มีการสร้างองค์ความรู้เพื่อเก็บเป็นข้อมูลคำศัพท์บางที่มาจากเว็บไซต์ที่สร้างขึ้นเพื่อรวบรวมคำศัพท์ที่เกี่ยวข้อง แต่อย่างไรก็ตามรูปแบบของการเรียนรู้แบบอัตโนมัติยังคงมีความจำเป็น



ภาพที่ 2.8 ตัวอย่างการนำคำศัพท์มาสร้างความสัมพันธ์ที่ถูกแก้ในฐานข้อมูลภาพ

<https://www.groundai.com/project/dynamic-structured-semantic-propagation-network6799/>

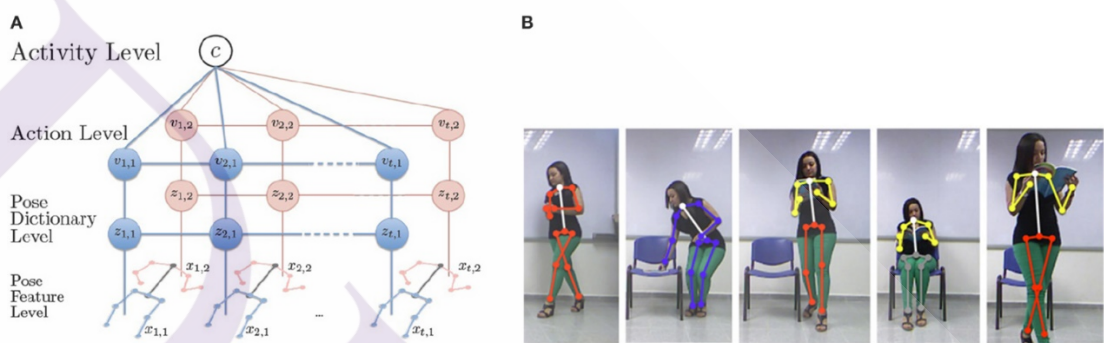
ปัจจุบันงานวิจัยพยายามที่จะสร้างความสัมพันธ์ภายในภาพ [J. Long 2015][Ren, S. 2015] เช่น (girl, on, horse) หรือ (man, eat, apple) เพื่อตอบสนองการแปลความหมายภาพที่มีความซับซ้อนขึ้น ทีมวิจัยของ Cewu Lu และคณะ [Bangpeng Yao 2010] ได้นำเสนอวิธีการที่ทำการจับคู่วัตถุ หัวเรื่องและความสัมพันธ์ระหว่างกัน ตัวอย่างเช่น รูปภาพที่ประกอบด้วย บุคคล (person), มอเตอร์ไซด์ (motorcycle) และ หมวกกันน็อค (helmet) สามารถสร้างความสัมพันธ์ได้เป็น (person - on - motorcycle), (person - wear - helmet) หรือ (motorcycle - has - wheel) ข้อมูลภาพทั้งหมดถูกรวบรวมและจำแนกด้วยวิธีการ Deep Relational Network จากผลการทดลองจะเห็นว่าการใช้รูปแบบความสัมพันธ์กับวัตถุนั้นภาพมีความหลากหลาย ทำให้การทำนายความสัมพันธ์ที่เกิดขึ้น จากการทดลองทั้งหมดต้องมีตัวอย่างภาพครอบคลุมมากเพียงพอถึงจะสามารถสื่อถึงความหมายของภาพโดยรวมอย่างแท้จริง

2.3 การค้นคืนด้วยกิจกรรมมนุษย์

งานวิจัยทางการประมวลผลภาพดิจิทัลเข้ามามีบทบาทในการช่วย การประมวลผลผลภาพดิจิทัลในการวิเคราะห์พฤติกรรมมนุษย์ และเฝ้าระวังเป็นหัวข้อวิจัยที่ยังคงเป็นหัวข้อวิจัยที่มีความท้าทาย สำหรับงานวิจัยที่พบอย่างแพร่หลายและมักจะคุ้นเคยในหัวข้อการรู้จำกิจกรรมมนุษย์ (Human activity recognition: HAR) [Michalis Vrigkas 2015] โดยมีจุดมุ่งหมายในการแปลความหมายของกิจกรรมของบุคคลหรือกลุ่มบุคคลด้วยพื้นฐานของกลุ่มข้อมูลที่ได้มาจากเซ็นเซอร์หรือวิดีโอ โดยส่วนใหญ่การวิเคราะห์กิจกรรมของมนุษย์เพื่อนำมาใช้ในการดำเนินการควบคุมกิจกรรมต่างๆของมนุษย์ เช่น การควบคุมความปลอดภัย (security control) กระบวนการเฝ้าระวัง (surveillance processes) การตรวจสอบการกิจกรรม (monitor activity) หรือกระบวนการดูแลสุขภาพผู้ป่วย (processes of monitoring a patients' health) เป็นต้น ปัจจัยหลักที่ทำให้กระบวนการตรวจสอบกิจกรรมมนุษย์เพิ่มพูนความสำคัญขึ้นอย่างต่อเนื่องนั้นคือ อุปกรณ์และกล้องดิจิทัลที่มีราคาถูกลง ประสิทธิภาพของการประมวลผลในระบบเวลาจริงที่มีความรวดเร็ว จึงทำให้การตอบรับของผู้ใช้งานเพิ่มมากขึ้น มีการแบ่งรูปแบบการพัฒนาวิธีการเพื่อทำการรู้จำกิจกรรม สามารถจัดกลุ่มได้เป็น 3 กลุ่มดังนี้

(1) Statistical Approaches เป็นการทำรู้จำกิจกรรมของมนุษย์ด้วยรูปแบบทางสถิติ เช่น Bayesian networks (BNs), hidden Markov models (HMMs) หรือ dynamic Bayesian networks (DBNs) [S. Fine, 1998][P. Natarajan, 2007] โดยทั่วไปแล้วรูปแบบทางสถิติหนึ่งรูปแบบจะสามารถสร้างแต่ละกิจกรรมได้ ด้วยการสร้างรูปแบบความน่าจะเป็นจากลำดับภาพที่ถูกละทิ้งออกมาจาก กิจกรรมที่มีความสัมพันธ์กันเป็นเวกเตอร์จากคุณสมบัติภาพ Park and Aggarwal [S. Park 2004][S.Park 2004] ได้แสดงรูปแบบลำดับขั้นที่ทำการรู้จำกิจกรรมมนุษย์ด้วยการใช้พิกเซลของภาพ ซึ่งจะแสดงเป็นท่าทางของมนุษย์ และกิจกรรม รวมทั้งการใช้ความสัมพันธ์ ซึ่งในระบบนี้เป็นการสกัดส่วนของร่างกายออกมาเพื่อทำการหาส่วนที่เด่นของท่าทางจากลำดับท่าทางทั้งหมดด้วย [Lillo, I. 2014] ดังแสดงในภาพที่ 2.8 แสดงตัวอย่างของการจำแนกลักษณะท่าทางของมนุษย์ด้วยการแทนของกราฟ การเรียนรู้จากการใช้

วิธีการ Bayesian networks จะเห็นว่าจากระบบที่มีการรู้จำกิจกรรมมนุษย์ด้วยการจำแนกจากโครงสร้างลำดับของกิจกรรมที่มีการสร้างความสัมพันธ์ของลำดับท่าทาง จากเหตุการณ์ย่อย เนื่องจากการทำงานมีลำดับของเหตุการณ์และท่าทางดังนั้น Nguyen et al. [N. T. Nguyen 2005] ได้นำรูปแบบลำดับชั้นของวิธีการ HMMs มาใช้ในการรู้จำท่าทางซึ่งจะมีการทำงานคล้ายกันกับงานของ Park and Aggarwal's [S. Park 2004] ซึ่งจะเห็นว่าความรู้จำสามารถทำได้ 2 ท่าทางจะเห็นว่าการใช้รูปแบบของสถิติเพื่อทำการรู้จำลำดับท่าทาง นั้นจะมีข้อจำกัดอยู่ที่ข้อมูลการเรียนรู้ และรูปแบบทางสถิติที่มีความสัมพันธ์กับการรู้จำท่าทางและข้อมูลที่มีความรัดกุมมากน้อยเพียงใดเป็นสิ่งที่จำเป็นอย่างมาก



ภาพที่ 2.9 ตัวอย่างการจำแนกท่าทางมนุษย์

<https://www.frontiersin.org/articles/10.3389/frobt.2015.00028/full>

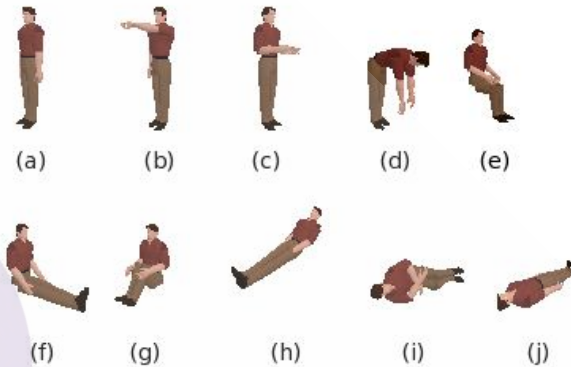
(2) Syntactic Approaches เป็นรูปแบบกิจกรรมที่ได้มาจาก Production rule ที่หลากหลายทั้งการสร้างสตริงสัญลักษณ์และการใช้ภาษาโปรแกรมในการรู้จำกิจกรรมถูกเรียกว่า วิธีการ Contextfree grammar (CFG) และวิธีการ Stochastic context-free grammar (SCFG) [A. F. Bobick 1997][D. Minnen 2003] สำหรับการรู้จำกิจกรรม Production rules ของ CFGs เป็นวิธีแรกเริ่มที่นำไปสู่การรู้จำกิจกรรมของมนุษย์ด้วยสัญลักษณ์ของสตริงจากการกระทำพื้นฐาน Ivanov and Bobick [Y. A. Ivanov 2008] ได้นำเสนอรูปแบบการรู้จำกิจกรรมโดยใช้ SCFG ซึ่งมีการใช้ HMM สำหรับการรู้จำลำดับท่าทางพื้นฐานด้วยสตริงและคำศัพท์ที่มีการดัดแปลงมา Minnen et al. [D. Minnen, 2003.] มีการใช้รูปแบบ SCFG สำหรับการรู้จำ โดยที่ระบบมีการแบ่งแยกส่วน (Segmentation) ของวัตถุและพยายามทำให้เข้าถึงการแปลความหมายของกิจกรรมด้วย CFG เพื่อช่วยในการแบ่งแยกส่วนวัตถุ แต่อย่างไรก็ตาม ยังมีข้อจำกัดของการรู้จำกิจกรรมย่อยด้วยการใช้วิธี Syntactic approaches เพราะเป็นการใช้ Production rules ของทุกเหตุการณ์เป็นโดเมนที่ใหญ่จึงไม่สามารถรู้จำได้ทุกกิจกรรมย่อย

(3) **Description-Based Approaches** เป็นวิธีการรู้จำกิจกรรมของมนุษย์ด้วยการใช้การแทนค่าทั้งในรูปแบบ โครงสร้าง temporal และ spatial ของกิจกรรม [S. Hongeng 2004][V.-T. Vu 2003] โดยที่จะมีการแทนความสัมพันธ์ในรูปแบบเทอมของ กิจกรรมและกิจกรรมย่อย สำหรับวิธีการของ Allen and Ferguson [J. F. Allen 1994] จะสร้างความสัมพันธ์ตามลำดับ sequential และความสัมพันธ์ที่เป็น concurrent และ ความสัมพันธ์ที่เกิดจากการ combinations แต่ความสัมพันธ์ที่เกิดจากการรวมกันจะมีความซับซ้อนกว่าจะมีการใช้วิธีการของ CFG เข้ามาช่วยในการแทนค่าความสัมพันธ์ในเชิงของภาษาโปรแกรมเพื่อแทนค่าและเขียนอัลกอริทึมในการรู้จำกิจกรรม Vu et al.'s [V.-T. Vu 2003] และ Allen and Ferguson [J. F. Allen 1994] ได้ทำการทดลองเพื่อแทนค่ากิจกรรมด้วยลำดับชั้น และนำไปประยุกต์ใช้เพื่อตรวจจับกิจกรรมการขโมยที่เกิดขึ้นในธนาคาร แต่ระบบของ Hongeng et al.'s [S. Hongeng 2004] ได้มีการเพิ่มตัวเชื่อมความสัมพันธ์ เช่น “and” และ “or” แต่ Siskind [J. M.Siskind 2001] ได้มีการเพิ่มช่วงเวลาและกิจกรรมย่อยลงได้ มีการใช้เหตุและผลกระทบความสัมพันธ์รวมทั้งการใช้ตัวเชื่อม “after” และโครงสร้างของการวนซ้ำเพิ่มเข้ามา

นักวิจัยพยายามพัฒนาวิธีการเพื่อทำการรู้จำกิจกรรม นำมาใช้กับการเฝ้าระวังจากกิจกรรมประจำวันของผู้สูงอายุเพื่อเป็นการป้องกันและดูแลผู้สูงอายุเพื่อ ทำการจำแนกความหมายของภาพกิจกรรมต่างๆจากข้อมูลวิดีโอ สำหรับนักวิจัยกลุ่มแรก ใช้ข้อมูลภาพสี privacy-violating RGB เพื่อทำการตรวจจับกิจกรรมพื้นฐานได้แก่ sitting, standing, walking, sleeping, using bedside commode, getting assistance เป็นต้น กิจกรรมเหล่านี้เป็นกิจกรรมประจำวันที่เกิดขึ้นกับมนุษย์ทั่วไป คณะที่วิจัยของ Haque [Albert Haque 2017] ได้นำภาพ RGB color และ คุณสมบัติของ texture เข้ามาใช้ในการตรวจสอบกิจกรรมการล้างมือภายในโรงพยาบาล แต่อย่างไรก็ตามท่าทางของมนุษย์ที่เกิดกิจกรรมโดยส่วนใหญ่มักเกิดการทับซ้อนกันของวัตถุ จึงเป็นปัญหาที่ทำให้เกิดความตีความที่ผิดพลาดของค่าความถูกต้อง อย่างไรก็ตามนักวิจัยหลายกลุ่มพยายามที่จะแก้ไขปัญหาก็เกิดขึ้นจากคุณสมบัติของภาพ และวิธีการต่างๆเพื่อเชื่อมโยงให้ได้ความหมายของภาพที่แสดงถึงความหมายของภาพที่สอดคล้องกับกิจกรรมอย่างแท้จริง



ก. การแบ่งส่วนวัตถุ เพื่อทำการจำแนกตัวมนุษย์



ข. การจำแนกท่าทางมนุษย์



ค. การรู้จำกิจกรรมของมนุษย์

ภาพที่ 2.10 แสดงตัวอย่างกระบวนการรู้จำกิจกรรมของมนุษย์

<http://www-sop.inria.fr/members/Francois.Bremond/topicsText/gerhomeProject.html>

โครงการ GER'HOME [Nadia Zouba 2007] เป็นการทำงานที่ตรวจดูกิจกรรมของผู้สูงอายุในชีวิตประจำวันด้วยการมอนิเตอร์ที่บ้านเพื่อลดค่าใช้จ่ายของโรงพยาบาลของการพยาบาลหลังจากเกิดอุบัติเหตุหรือลดอาการบาดเจ็บของผู้ป่วยได้ก่อน ด้วยการติดตั้งกล้องและเซ็นเซอร์ตามจุดต่างๆในบริเวณบ้าน ดังแสดงในภาพที่ 2.9 ค. และจากภาพที่ได้จะทำการแบ่งแยกวัตถุดังแสดงในภาพที่ 2.9 ก. เพื่อนำวัตถุมารเปรียบเทียบเป็นแพทเทิร์นด้วยการใช้รูปแบบ 3D ที่ถูกกำหนดเป็นท่าทางไว้เพื่อที่จะรู้จำท่าทางมนุษย์ [V. Joumier 2012][C.

Crispim-Junior 2013] ดังแสดงในภาพที่ 2.9 ข. โดยกิจกรรมทั่วไป จะประกอบด้วย กิจกรรมทั่วไป walking, standing, sitting, lying, Bending และกิจกรรมที่ต้องแฉ่งเตือนคือ fainting และ falling down [B. Fosty 2015] ด้วยวิธีการ Hidden Markov Model (HMM) และ Neuronal Networks (NNs) เป็นลำดับในการวิเคราะห์ท่าทางที่เกิดเหตุการณ์ขึ้นแต่อย่างไรก็ตามการวิเคราะห์ด้วยวิธีการนี้ยังคงเกิดปัญหาสำหรับการรู้จำกิจกรรมกับบริเวณที่ควรสนใจจึงมีการใช้อัลกอริทึม K-means clustering [F. Negin 2014] เพื่อจำกัดจุดบริเวณที่สนใจกลายเป็นจำนวนคสเตอร์ที่กำหนดไว้เพื่อให้ง่ายกับการจำแนกท่าทางต่อไป

จากกระบวนการวิจัยในส่วนของการค้นคืนกิจกรรมมนุษย์เพื่อมอนิเตอร์ผู้สูงอายุส่วนใหญ่เป็นการใช้รูปแบบวิดีโอและกระบวนการประมวลผลระดับต่ำเพื่อที่จะรู้จำและตัดสินใจในกิจกรรมที่เกิดขึ้นทันที ซึ่งยังต้องใช้เวลาและการตีความ ดังนั้นในงานวิจัยนำเสนอวิธีการจำแนกข้อมูลแบบใหม่ด้วยการสร้างพารามิเตอร์ของความสัมพันธ์ระหว่างวัตถุและแอคชันจากข้อมูลภายในภาพเพื่อสามารถนำไปใช้ในการหาภาพผลลัพธ์ที่ตรงความต้องการของผู้ใช้งานได้มากขึ้น ด้วยการสร้างความสัมพันธ์ภายในภาพจากความรู้พื้นฐาน เช่น (girl, on, horse) หรือ (man, eat, apple) เพื่อตอบสนองการแปลความหมายภาพที่มีความซับซ้อนโดยเฉพาะการแปลความหมายเพื่อช่วยเหลือผู้สูงอายุ ด้วยการที่ทำการจับคู่วัตถุ หัวเรื่องและความสัมพันธ์ระหว่างกัน ภายในรูปภาพที่ประกอบด้วยวัตถุต่างๆ เช่น บุคคล (person), เครื่องพุงเดิน (walker) และ ไม้เท้า (stick) สามารถสร้างความสัมพันธ์ได้เป็น (person - hold - walker), (person -hold - stick) ดังนั้นในงานวิจัยนี้จึงได้นำเสนอวิธีการทำนายจำแนกข้อมูลด้วยวิธีการแบบใหม่ ด้วยการสร้างพารามิเตอร์ของความสัมพันธ์ระหว่างวัตถุและแอคชันจากข้อมูลภายในภาพ และคำศัพท์ที่ถูกนำมาใช้จะมีการเชื่อมโยงด้วย WordNet ที่เป็นข้อมูลความสัมพันธ์ของคำศัพท์แบบเป็นหมวดหมู่ ด้วยวิธีการทำนายความสัมพันธ์ด้วยการปรับค่าน้ำหนัก เพื่อสามารถนำไปใช้ในการหาภาพผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้งานได้มากขึ้น ขั้นตอนการจำแนกความหมายภาพกิจกรรม

สำหรับในงานวิจัยนี้ได้พัฒนาวิธีการเพื่อให้สามารถจำแนกความหมายภาพให้ได้ตรงกับสิ่งที่ผู้ใช้งานต้องการมากที่สุด จึงได้นำเสนอรูปแบบของการเชื่อมโยงความสัมพันธ์ระหว่างวัตถุและแอคชันจากข้อมูลภาพแบ่งลำดับขั้นตอนวิธีการเป็น 4 ส่วนหลักดังนี้ (1) ขั้นตอนการเตรียมข้อมูล (Data preprocessing) (2) การตรวจจับแอคชันบนภาพ (Action Detection) (3) การทำนายความสัมพันธ์แอคชัน (Predict relations of actions) (4) การวัดและการประเมินผลการทำงาน (Measurement and Evaluation)

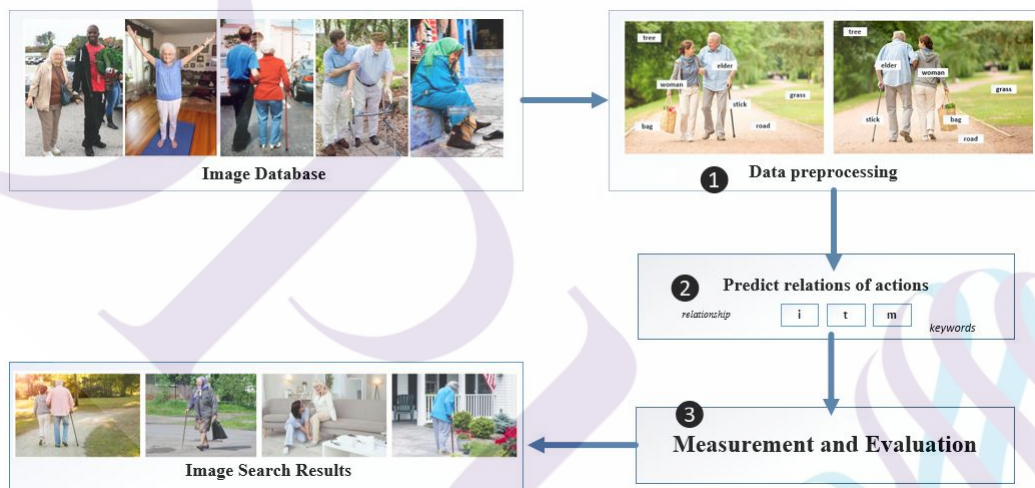
บทที่ 3 การดำเนินงานวิจัย

การศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยโดยทั่วไปจะแบ่งออกเป็น 3 ส่วนหลักดังนี้

3.1 การเตรียมข้อมูล (data preprocessing) เป็นการทำงานในส่วนของการนำข้อมูลเข้าด้วยเครื่องมือ และการแทนวัตถุลงในกราฟ

3.2 การทำนายความสัมพันธ์แอ็คชัน (predict relations of actions) เป็นการทำนายคาดการณ์ความสัมพันธ์ที่เกิดขึ้นภายในภาพ

3.3 การวัดและการประเมินผลการทำงาน (measurement and evaluation) เป็นการเปรียบเทียบการทำงานของวิธีการที่นำเสนอโดยขั้นตอนทั้งหมดจะสามารถอธิบายได้ดังภาพที่ 3.1



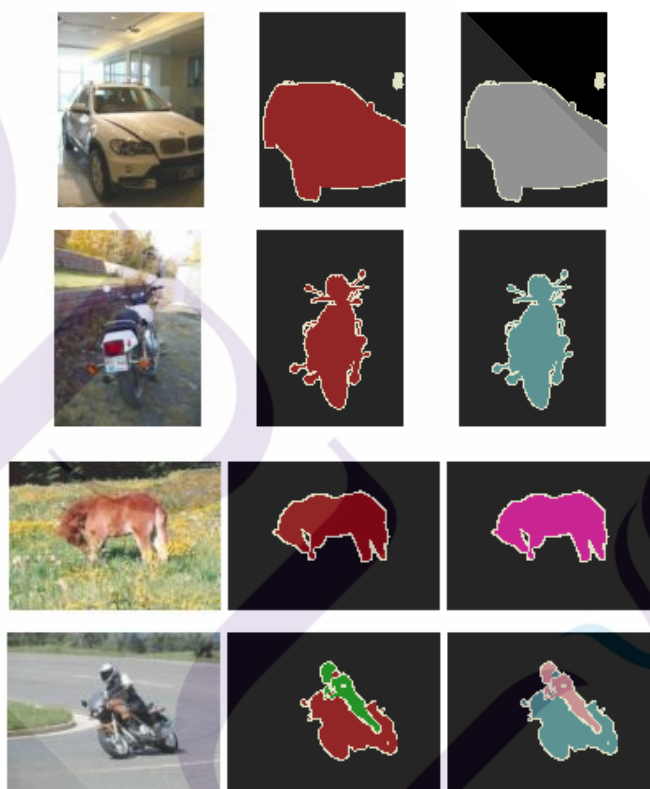
ภาพที่ 3.1 ขั้นตอนการจำแนกความหมายภาพด้วยการทำนายความสัมพันธ์คำศัพท์บนภาพ

3.1 การเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการทำการแยกและคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ สำหรับภาพบางภาพจะไม่นำมาทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวม ภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโพกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อทำให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้

3.1.1 ฐานข้อมูลรูปภาพ

แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพได้ เช่น Pascal Fotosearch stock [Everingham, M 2010], The Cobis Stock, The Corel Corporation เป็นต้น มีกลุ่มงานวิจัยหลายกลุ่มที่พยายามปรับปรุงวิธีการเพื่อทำการตรวจจับวัตถุ (object recognition) และตรวจจับแอ็คชันบนภาพ (action recognition) ผลการทดลองที่ได้จากการตรวจจับวัตถุ และแอ็คชันมีค่าความถูกต้องค่อนข้างดีและสามารถนำวิธีการมาประยุกต์ใช้ได้ ดังนั้นในงานวิจัยนี้ได้เจาะจงนำเสนอวิธีการใหม่เพื่อหาความสัมพันธ์ของวัตถุและแอ็คชัน จึงได้เลือกวิธีการที่ตรวจจับวัตถุและแอ็คชันบนภาพ พร้อมเลือกฐานข้อมูลที่มีขนาดไม่ใหญ่มาก เพื่อทดสอบการทำงานที่ครอบคลุม จึงเลือกฐานข้อมูลแอ็คชันจาก PASCAL2 [Everingham, M 2010] Activitynet [F. Caba Heilbron 2015] และ ImageNet [J. Deng, 2009] เพื่อเป็นพื้นฐานของการทดลองตามรูปแบบของแอ็คชัน ที่กำหนดไว้



ก. ภาพ

ข. วัตถุ

ค. คลาส

ภาพที่ 3.2 ตัวอย่างการแบ่งส่วนในฐานข้อมูล Pascal2

<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/segexamples/index.html>

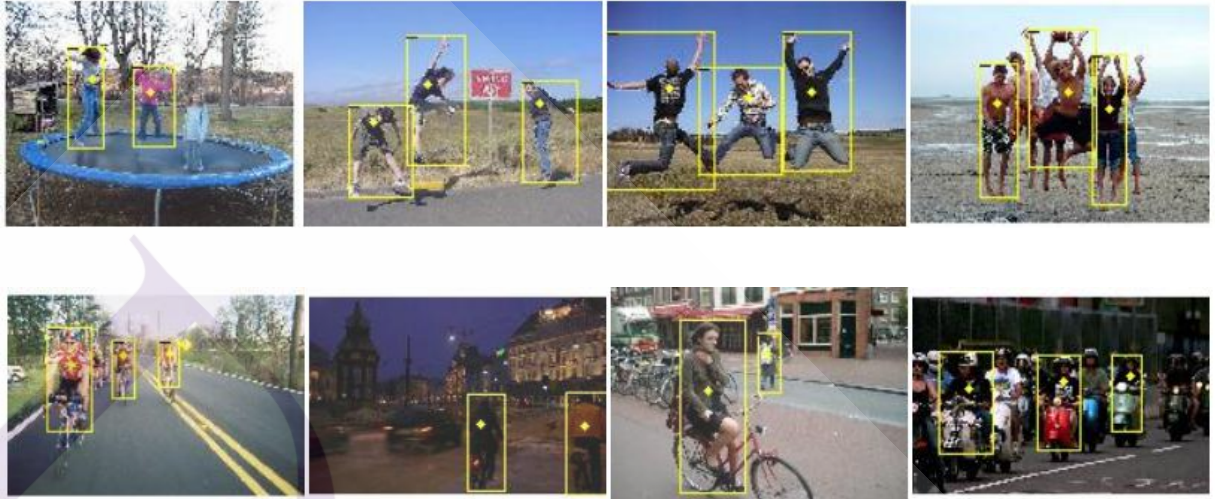
Pascal2 (Visual Object Classes Challenge 2012: VOC2012) [Everingham, M 2010] เป็นทีมงานวิจัยที่ทำงานสานต่อมาจาก โปรเจค VOC โดยใช้วิธีการจาก VOC2012 และได้ถูกพัฒนาขึ้นอย่างต่อเนื่องจนสามารถประกาศเป็นสาธารณะ บนออนไลน์ได้ และยังได้มีการร่วมกับ ECCV2019 และ ImageNet ในปี 2012 การทำงานของ Pascal2 พยายามจะสกัดข้อมูลภาพที่ได้เป็นชื่อของวัตถุที่ถูกกำหนดไว้ซึ่งในงานวิจัยได้มีการทดลองถึงวัตถุหลายอย่าง แบ่งเป็นกลุ่มได้ 20 กลุ่ม ตัวอย่างเช่น

Person: person

Animal: bird, cat, cow, dog, horse, sheep

Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train

Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor



ก. jumping

ข. ridding bike

ค. Walking

ง. Running

ภาพที่ 3.3 ตัวอย่างการจำแนกแอ็คชันบนฐานข้อมูล Pascal2



<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/actionexamples/index.html#running>

กลุ่มการแบ่งส่วนเพื่อนำพิกเซลลงกลุ่มที่มีความหมาย ประกอบด้วย (1=aeroplane, 2=bicycle, 3=bird, 4=boat, 5=bottle, 6=bus, 7=car , 8=cat, 9=chair, 10=cow, 11=diningtable, 12=dog, 13=horse, 14=motorbike, 15=person, 16=potted plant, 17=sheep, 18=sofa, 19=train, 20=tv/monitor) ดังแสดงอย่างกลุ่มในภาพที่ 3.2 สำหรับแอ็คชันที่มีการจำแนกและรู้จำวัตถุในฐานข้อมูล Pascal2 ในแต่ละกล่อง ตัวอย่างเช่น Jumping, Phoning, PlayingInstrument, Reading, RidingBike, RidingHorse, Running, TakingPhoto, UsingComputer, Walking ดังแสดงในภาพที่ 3.3 สำหรับแหล่ง

ข้อมูลภาพที่หลากหลาย อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอเนื่องจาก ภาพบางภาพมีลักษณะ ผิดปกติหรือ คุณลักษณะวัตถุไม่ชัดคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้ชื่อวัตถุได้ ภาพถ่าย ระยะใกล้ ภาพบางภาพอาจไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถ หาความหมายภาพได้ ทำให้ต้องมีการคัดเลือกภาพออกไป ไม่นำมาใช้ในการทดลอง ดังนั้นในการหา แหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ภาพจะต้องความเหมาะสมกับงานที่จะนำมาใช้ เพื่อตอบสนองกับความต้องการของการทดลองมากที่สุด

สำหรับการทดลองโดยทั่วไปในสาขาคอมพิวเตอร์วิชัน (computer vision) ในส่วนของการประมวลผล ภาพระดับสูง รูปภาพที่นำเข้ามาทดลองจะเป็นภาพที่วัตถุถูกแท็ก หรือกระทำการ ให้ความหมายมาล่วงหน้า ก่อน จะเรียกภาพจำพวกนี้ว่า annotated images [J. Long 2015] ดังแสดงในภาพที่ 3.4 แสดงรูปภาพถูก แท็กด้วยคำหลัก ดังนั้นในการทดลองจะต้องมีฐานข้อมูลภาพที่สมบูรณ์เพียงพอที่จะสามารถนำส่วนของคำหลัก ที่ถูกแท็กมาใช้งานได้โดยไม่มีผลข้างเคียงต่อกระบวนการที่นำเสนอ ส่วนใหญ่แหล่งข้อมูลภาพจะทำการคัดเลือก บริเวณ ที่เหมาะสมสำหรับการแท็กเป็นคำหลักหรือคำสำคัญ เพื่อใช้สำหรับการสืบค้นข้อมูล ดังนั้นบาง แหล่งข้อมูลจะใช้การแบ่งหมวดหมู่วัตถุ (object categories) [Sun Ting 2016] เข้ามาช่วยในการจัดกลุ่มของ คำที่จะนำมาทดแทนส่วน บริเวณนั้นๆ

{act, action, activity}	{natural object}
{animal, fauna}	{natural phenomenon}
{artifact}	{person, human being}
{attribute, property}	{plant, flora}
{body, corpus}	{possession}
{cognition, knowledge}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation}
{feeling, emotion}	{shape}
{food}	{state, condition}
{group, collection}	{substance}
{location, place}	{time}
{motive}	

ภาพที่ 3.4 รายการคำหลักเริ่มต้นบน WordNet

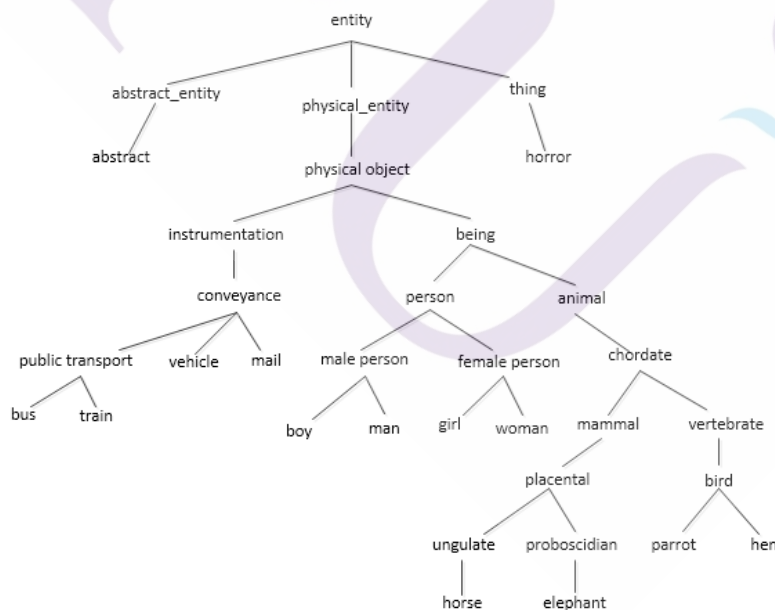
3.1.2 ฐานข้อมูลคำหลัก WordNet

WordNet [George A. Miller 1990] [Zinger S., 2005] เป็นฐานข้อมูลคำศัพท์ภาษาอังกฤษขนาดใหญ่ของ คำนามคำกริยา คำคุณศัพท์ และ กริยาวิเศษณ์ มีการจัดกลุ่มเป็นชุดของคำพ้องความรู้ความเข้าใจ (synsets) แต่ละการแสดงแนวคิดที่แตกต่างกัน synsets จะเชื่อมโยงกันโดยวิธีการของความสัมพันธ์ของ แนวคิดและความหมายคำศัพท์ เครือข่ายที่เกิดจากคำที่เกี่ยวข้องกับความหมายและแนวคิดสามารถสำรวจที่มี เบลาร์เซอร์ WordNet ยังเป็นอิสระและเปิดเผยต่อสาธารณชนพร้อมสำหรับการดาวน์โหลด โครงสร้างมโนทัศน์ ทำให้เครื่องมือที่มีประโยชน์สำหรับภาษาศาสตร์คอมพิวเตอร์และการประมวลผลภาษาธรรมชาติ

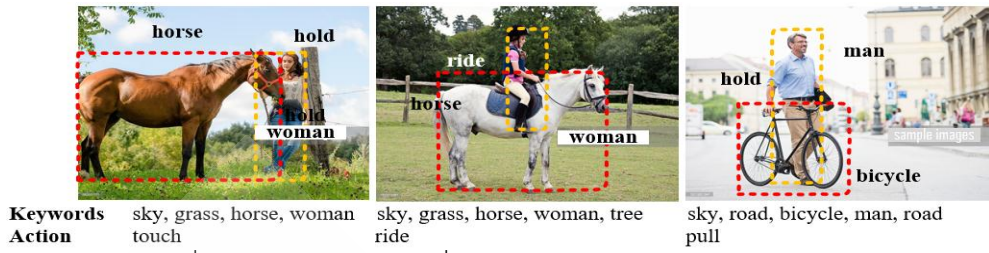
WordNet เป็นกลุ่มคำที่มันขึ้นอยู่ร่วมกันตามความหมาย WordNet ดังแสดงในภาพที่ 3.4 เป็น ฐานข้อมูลเกี่ยวกับคำศัพท์ที่ใช้ภาษาอังกฤษที่ประกอบด้วยคำนาม คำกริยา คำคุณศัพท์ และคำต่างๆถูกจัดไว้ เป็นหมวดหมู่ซึ่งเป็นงานวิจัยที่ได้รับความนิยมมากมีนักวิจัยหลายกลุ่มที่อ้างอิงการใช้ฐานข้อมูลของคำหลักที่

แท็กบนภาพ โดย WordNet จะใช้คำนาม ถูกเลือกจากพจนานุกรมเป็นคำหลัก โดยทั่วไปแล้วมาตรฐานอักษรจะมีการจัดการข้อมูล lexical เพื่อให้มีความสอดคล้องกันทั้งการสะกดคำ การกระจายความหมายเหมือนของคำ ความสัมพันธ์ของคำ มีการสร้างฐานข้อมูลคำศัพท์ออนไลน์ด้วยคอมพิวเตอร์ที่เรียกว่า lexical database เพื่อค้นหาแต่ละตัวอักษรได้อย่างรวดเร็ว WordNet เป็นส่วนของโปรแกรมที่พยายามรวมข้อมูล lexicographic ที่มีการคำนวณอย่างรวดเร็ว ซึ่งจะมีการจัดการข้อมูล lexical ในเทอมของความหมายจนถึงรูปแบบของคำศัพท์โดยจะใช้ Lexical Matrix ช่วยในการจัดการการกวैयाกรณ์ของคำและลตรูปแบบของคำ คำกำกวม ความหมายของคำ (word meaning) โดยจะใช้รูปแบบของการ mapping ระหว่าง form และ ความหมาย [George A. Miller 1993] ซึ่งจะสนใจเรื่องของกลุ่มกวैयाกรณ์คำโดยจะใช้การ mapping ที่มีรูปแบบต่างกัน

ในขั้นตอนทั้งหมดสิ่งสำคัญที่ช่วยทำให้กระบวนการสร้างการเชื่อมโยงความสัมพันธ์ของความหมายภาพให้มีความสมบูรณ์ขึ้นเพื่อเข้าสู่กระบวนการเรียนรู้อย่างสมบูรณ์เต็มรูปแบบ คือขั้นตอนการเตรียมข้อมูล เริ่มจากคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพเด่นชัด มีวัตถุภาพพื้นหลังชัดเจน ส่วนภาพมีลักษณะผิดปกติหรือ คุณลักษณะวัตถุไม่ชัดเจนคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถบ่งชี้วัตถุได้หรือภาพถ่ายระยะใกล้ จะถูกคัดออก ดังนั้นภาพที่คัดเลือกเข้ามาจะต้องมีความสมบูรณ์ของวัตถุและคำศัพท์ชัดเจน และแปลความหมายภาพนั้นได้อย่างสมบูรณ์ ความหมายของคำหลักที่ถูกนำมาใช้บน WordNet นั้นถูกจัดลงบนลำดับชั้นตามความหมายที่สามารถเป็นไปได้โดยที่ WordNet ได้กำหนดรายการหัวข้อหลักต่างๆตั้งต้นไว้ดังแสดงในภาพที่ 3.4 เป็นกลุ่มหัวข้อเริ่มต้นเป็นลำดับบนสุดของกราฟจะเป็น {entity} {object, thing} และ {idea} เป็นหัวข้อเริ่มต้นดังนั้นคำศัพท์ที่ถูกใช้จะอยู่ในหัวข้อต่างๆที่กำหนดซึ่งบางคำมีหลายความหมายสามารถอยู่ได้ในกลุ่มหลายกลุ่มตามความหมายที่เกิดขึ้น

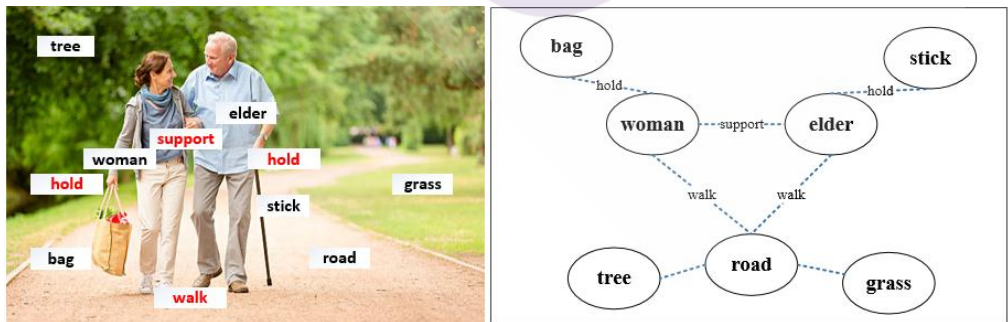


ภาพที่ 3.5 ตัวอย่างจำลอง Synset ในอนุกรมวิธาน WordNet



ภาพที่ 3.6 ตัวอย่างคำศัพท์ที่ประกอบด้วยคำนามและแอ็คชันบนฐานข้อมูลภาพ

ข้อมูลแอ็คชันบนภาพ เกิดจากความสัมพันธ์ของการแอ็คชันของคำนามที่เกิดขึ้น รวมทั้งการลดและจำกัดรูปแบบความสัมพันธ์ เพื่อให้เกิดความรัดกุมของข้อมูลเป็นอีกสิ่งจำเป็นเพื่อให้ภาพเกิดความหมายเดียวกันมากขึ้น สำหรับแต่ละคำศัพท์การแอ็คชันกำหนดให้มีรูปแบบที่ประกอบด้วยโครงสร้าง คำศัพท์ที่ทำหน้าที่เป็นประธาน หรือคำศัพท์ที่เป็นคำนาม และ คำศัพท์ที่เป็นแอ็คชัน (กริยา) ตัวอย่างดังแสดงในภาพที่ 3.6 “woman holding horse”, “woman riding on horse” หรือ “man holding bike” เป็นต้น วัตถุที่ปรากฏบนภาพ woman horse man และbike มีแอ็คชัน holding และ riding เป็นต้น ดังนั้นในงานวิจัยนี้ ภาพจะคัดเลือกเข้ามามีความสมบูรณ์ของวัตถุและคำศัพท์ชัดเจนตามกลุ่มที่ต้องการ สำหรับข้อมูลกิจกรรมบนภาพ ได้ใช้วิธีการตรวจจับท่าทางในแต่ละกลุ่มมาจาก Activitynet [F. Caba Heilbron 2015] กระบวนการประมวลผลและการระบุวัตถุและภาพจะมีความสัมพันธ์กับกิจกรรมที่เกิดขึ้น สำหรับความสัมพันธ์ที่เกิดขึ้นระหว่างวัตถุ เช่น ท่าทางกับวัตถุ และ ท่าทางกับภาพพื้นหลังจะมีการกำหนดด้วยมือซึ่งจะใช้คำศัพท์จากกลุ่ม ImageNet [J. Deng, 2009] ดังนั้นจะมีการจำกัดรูปแบบความสัมพันธ์ระหว่างวัตถุ เพื่อให้เกิดความรัดกุมของข้อมูลเป็นอีกสิ่งจำเป็นเพื่อให้ภาพเกิดความหมายเดียวกันมากขึ้น สำหรับแต่ละคำศัพท์การกิจกรรมกำหนดให้มีรูปแบบที่ประกอบด้วยโครงสร้าง คำศัพท์ที่ทำหน้าที่เป็นประธาน หรือคำศัพท์ที่เป็นคำนาม และ คำศัพท์ที่เป็นท่าทาง (กริยา) ดังแสดงตัวอย่างในภาพที่ 2 “woman supporting elder”, “woman walking on road” หรือ “elder holding stick” เป็นต้น วัตถุที่ปรากฏบนภาพ woman elder stick และ bag มีแอ็คชัน holding และ supporting เป็นต้น มีกลุ่มงานวิจัยหลายกลุ่มที่พยายามปรับปรุงวิธีการเพื่อทำการตรวจจับวัตถุ และตรวจจับกิจกรรมบนภาพ ผลการทดลองที่ได้จากการตรวจจับวัตถุและแอ็คชันมีความถูกต้องค่อนข้างดีและสามารถนำวิธีการมาประยุกต์ใช้ได้



ภาพที่ 3.7 ความสัมพันธ์ระหว่างวัตถุและท่าทางภายในภาพ

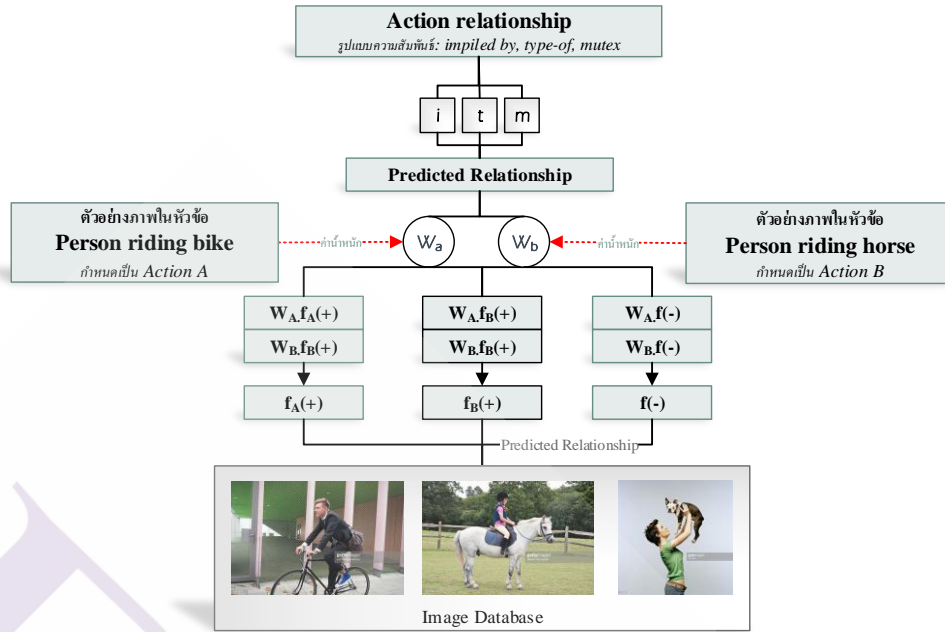
การแทนค่าข้อมูลวัตถุบนภาพลงในโมเดลพื้นฐานกราฟ โดยกำหนดให้ G เป็นกราฟที่ประกอบด้วยข้อมูล $G = \{(s, p, o)\}_{i=1}^N \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ เมื่อกำหนดให้ข้อมูลภายในประกอบด้วย $s, o \in \mathcal{E}$ แทนค่าเป็นหัวเรื่องและวัตถุของกราฟ โดยมีความสัมพันธ์ระหว่างสองสิ่งคือ $p \in \mathcal{R}$ การสร้างการเชื่อมต่อเพื่อทำนายสามารถเขียนอธิบายได้ด้วย $\theta: \mathcal{E} \times \mathcal{R} \times \mathcal{E} \rightarrow \mathbb{R}$ สามารถเขียนเป็น (s, p, o) เพื่อทำการแมพกลับไปเป็นค่าของข้อมูลจริงได้ แต่อย่างไรก็ตามการนำข้อมูลทั้งหมดมาใช้ในการทำนายนั้นมักจะใช้รูปแบบของการเรียนรู้จากการแทนค่าของข้อมูลที่มีความสัมพันธ์กัน ซึ่งส่วนใหญ่่มักจะใช้การทำนายความสัมพันธ์แอ็คชันสำหรับในงานนี้ได้เลือกวิธีการทำนายไว้ดังนี้

3.2 การทำนายความสัมพันธ์แอ็คชัน

ข้อมูลคำศัพท์ภาพ จะถูกแบ่งชนิดเป็น 2 ชนิดดังนี้ คำศัพท์ที่เป็นคำนาม (Noun) และคำศัพท์ที่เป็นแอ็คชัน (action) กำหนดให้ คำศัพท์แอ็คชัน (Action: A) เขียนความสัมพันธ์ของแอ็คชันบนภาพได้ I_A และความสัมพันธ์ระหว่างคำศัพท์การแอ็คชันกำหนดให้เป็น $R_A \subset A$ สำหรับทุกการแอ็คชัน ซึ่งมีค่าน้ำหนักของแอ็คชัน (weight of $A : \omega_A$) เป็น $\omega_A \in \mathcal{R}$ และแต่ละภาพ I จะมีข้อมูลคุณลักษณะภาพที่แทนด้วย $f_I \in \mathcal{R}$ โดยที่กำหนดข้อมูลให้ $f_I = W_{im} CNN(I) + b_{im}$, ถูกจัดเก็บข้อมูลในรูปแบบของลำดับชั้นโครงสร้างของโครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolution Neural Network: CNN) [Joachims, T. 1998] [Y. Tsuruoka, 2009] เมื่อกำหนดให้ $CNN(I)$ แทนคุณลักษณะของภาพ I และ กำหนดให้ พารามิเตอร์ W_{im} และ b_{im} เป็นข้อมูลที่ได้จากการเรียนรู้ระหว่างโมเดลการทำนายความสัมพันธ์ ดังนั้นสามารถกำหนดให้ ค่าน้ำหนักของแอ็คชันสูงสุดสำหรับภาพที่อยู่ในกลุ่มความหมายเดียวกัน เหมือนกัน คล้ายกันจะมีค่าเป็นค่าบวก (Positive image) ตัวอย่างเช่น ภาพคนขี่จักรยานที่มีท่าทางต่างๆดังแสดงในภาพที่ 3.3 ข. กลุ่มเดียวกัน หรือที่มีท่าทางการวิ่งในรูปแบบเดียวกันจะมีค่าเป็นบวก ดังแสดงในภาพที่ 3.3 ง. เป็นต้น และจะมีการเปรียบเทียบกับค่าที่เป็นลบ ดังนั้นทำให้เกิดการจัดลำดับของการเปรียบเทียบขึ้นและสามารถเขียนเป็นสมการของการจัดลำดับ (C_{ac}) ได้ เป็น

$$C_{ac} = \sum_A \sum_{\substack{I^+ \in I_A \\ I^- \in I_A}} \max(0, 1 + \omega_A^T (f_{I^-} - f_{I^+})) \quad (1)$$

เมื่อกำหนดให้ ข้อมูลคุณลักษณะภาพ (f_I) แทนภาพความหมายเดียวกันเป็น (I^+) และกำหนดให้ คุณลักษณะภาพที่มีความหมายต่างกัน (I^-) แทนด้วยสมการ $f_I = (f_{I^-}, f_{I^+}) \in \mathcal{R}$



ภาพที่ 3.8 ขั้นตอนการสร้างความสัมพันธ์ของแอ็คชัน

การทำนายความสัมพันธ์ของภาพเป็นการนำความสัมพันธ์ระหว่างวัตถุและแอ็คชันที่เกิดขึ้นบนภาพ มาสร้างความสัมพันธ์ภายใน โดยที่กำหนดคู่ความสัมพันธ์ของคำศัพท์การแอ็คชัน $A, B \subset R_A$ และกำหนดรูปแบบของความสัมพันธ์ที่เกิดระหว่างคำศัพท์ในภาพดังนี้

(1) แบบ implied-by คือ A implied – by B ถ้าเกิดแอ็คชันของ B แสดงให้เห็นว่าสามารถเกิดการแอ็คชัน A ได้ จะมีลักษณะของความสัมพันธ์แบบเดียวกับต้นไม้ กิ่งพ่อแม่ไปยังกิ่งลูก (parent-child node)

(2) แบบ type-of คือ A type – of B แอ็คชันของ A เป็นชนิดหนึ่งของการเกิดการแอ็คชัน B ได้จะมีลักษณะของความสัมพันธ์แบบเดียวกับต้นไม้ กิ่งลูกไปยังกิ่งพ่อแม่ (child-parent node)

(3) แบบ mutually exclusive เมื่อมีการแอ็คชันของ A เกิดขึ้นแล้วจะไม่เกิดขึ้นอย่าง B อย่างแน่นอน แสดงความสัมพันธ์ในภาพที่ 3.9 เป็นการทำนายความสัมพันธ์จากการเรียนรู้รูปแบบของแอ็คชันที่เกิดขึ้น (ภาพตัวอย่างมีแอ็คชัน: Action A และ Action B) จะเป็นค่าน้ำหนักที่ได้จากความสัมพันธ์ (กลายเป็นค่าน้ำหนักที่เกิดจากแอ็คชัน A: ω_A และแอ็คชัน B: ω_B) และจะเรียนรู้ตามรูปแบบความสัมพันธ์ทั้ง 3 แบบสามารถเขียนเป็นสมการความสัมพันธ์ได้ดังนี้ $r_{AB} = [r_{AB}^i, r_{AB}^t, r_{AB}^m] \in [0,3]^3$, เมื่อ r^i, r^t, r^m เป็นความสัมพันธ์ของ Implied-by, Type-of และ Mutually exclusive ตามลำดับ โดยที่ความสัมพันธ์ที่เกิดขึ้นจะใช้โครงข่ายประสาทเทียม (Neural network) ในการจัดการความรู้ ด้วยการจัดลำดับชั้นด้วย Softmax Normalization [Goodfellow, Ian 2016] สามารถเขียนเป็นสมการการคาดคะเนความสัมพันธ์ได้ดังนี้

$$r_{AB} = \text{softmax}_\beta(\omega_A \otimes W_\delta^{[1:3]} \otimes \omega_B + b_\delta), \quad (2)$$

เมื่อกำหนดให้ $W_{\delta}^{[1:3]} \in \mathbb{R}^{n \times n \times 3}$, n แทนจำนวนข้อมูลของภาพและ $b_{\delta} \in \mathbb{R}^3$

$\text{softmax}_{\beta}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ มีตัวแปร Softmax Normalization เป็น β สำหรับการสร้างความสัมพันธ์ของการแอ็คชัน (Create Action Relationships) เริ่มต้นจากการกำหนดให้สร้างความสัมพันธ์เพื่อใช้ในการเรียนรู้สำหรับการทำนายด้วยคำศัพท์ที่ใช้ในเซตความสัมพันธ์ กำหนดให้ความสัมพันธ์ระหว่างการแอ็คชันของวัตถุมีทั้งหมด 3 ประเภทดังนี้

(1) แบบ *A implied – by B* เมื่อเวกเตอร์ค่าน้ำหนักการแอ็คชัน (Action Weight Vector: W_A) เป็นอยู่ในลำดับภาพที่อยู่ในกลุ่มภาพ B ที่มีค่ามากกว่าภาพ A ที่ไม่อยู่ในกลุ่มแล้ว กำหนดให้มีค่าความคาดหวังความสัมพันธ์เป็น

$$C_{AB}^i = \sum_{I^b \in I_B, I^- \in I_A} \max(0, 1 + w_A^T(f_{I^-} - f_{I^b})) \quad (3)$$

(2) แบบ *A type – of B* เมื่อเวกเตอร์ค่าน้ำหนักการแอ็คชัน (Action Weight Vector: W_B) อยู่ในลำดับภาพที่อยู่ในกลุ่มภาพ A ที่มีค่ามากกว่าภาพ B ที่ไม่อยู่ในกลุ่มแล้ว กำหนดให้มีค่าความคาดหวังความสัมพันธ์เป็น

$$C_{AB}^t = \sum_{I^b \in I_A, I^- \in I_B} \max(0, 1 + w_B^T(f_{I^-} - f_{I^a})) \quad (4)$$

(3) แบบ *A mutually exclusive B* เมื่อเวกเตอร์ค่าน้ำหนักการแอ็คชันเป็น W_A อยู่ในลำดับภาพที่อยู่ในกลุ่มภาพ A ที่มีค่ามากกว่าภาพ B ที่อยู่ในกลุ่มแล้วจะกำหนดให้มีค่าความคาดหวังความสัมพันธ์เป็น

$$C_{AB}^m = \sum_{I^a \in I_A, I^- \in I_B} \max(0, 1 + w_A^T(f_{I^b} - f_{I^a})) \quad (5)$$

ดังนั้นทำการรวมสมการ (3) ถึง (5) เข้าด้วยการเพื่อสร้างความสัมพันธ์ของการคาดทำนาย ดังแสดงในสมการ

$$C_{\alpha} = \sum_{A \in \mathcal{A}, B \in \mathcal{R}_A} r_{AB}^i \cdot C_{AB}^i + r_{AB}^t \cdot C_{AB}^t + r_{AB}^m \cdot C_{AB}^m \quad (6)$$

เวกเตอร์ค่าน้ำหนักการแอ็คชัน W_A และ W_B ได้มาจากการเรียนรู้ของฟังก์ชันความสัมพันธ์ที่ดีที่สุดของข้อมูลภาพ ทำให้วิธีการสามารถเลือกความสัมพันธ์ได้อย่างอัตโนมัติเพื่อทำการคำนวณค่าน้ำหนักของการแอ็คชันได้ ดังแสดงตัวอย่างความสัมพันธ์ในภาพที่ 3.9



ภาพที่ 3.9 ตัวอย่างเซตความสัมพันธ์ของการแอ็คชันของข้อมูลภาพ

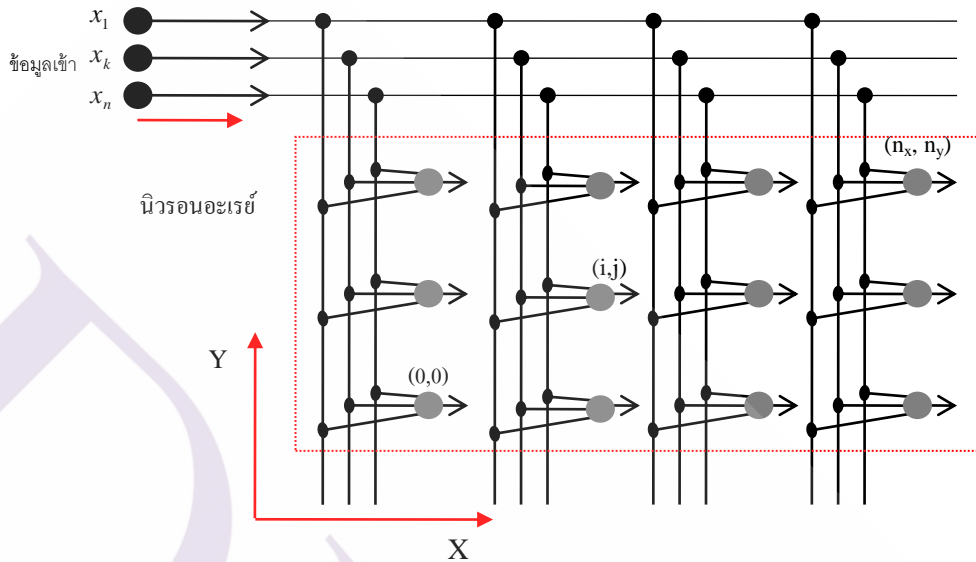
3.3 การวัดและการประเมินผลการทำงาน

การวัดและประเมินผลการทำงาน (Measurement and Evaluation) เป็นขั้นตอนสุดท้ายที่สำคัญเพื่อทดสอบวิธีการข้างต้นว่ามีประสิทธิภาพมากหรือน้อย สอดคล้องกับความต้องการเพียงใดเมื่อนำมาใช้งานจริง สำหรับข้อมูลภาพที่ใช้ในการทดลองทั้งหมดจะคัดเลือกภาพที่เห็นวัตถุอย่างเด่นชัดและมีความสอดคล้องกับฐานข้อมูลพื้นฐานจากฐานข้อมูลภาพแอ็คชัน Pascal [Everingham, M.2010] โดยที่แต่ละกลุ่มคำจะมีแอ็คชันที่มีความสอดคล้องกันทั้งภาพที่อยู่ในกลุ่ม (Positive Image) และภาพที่ไม่อยู่ในกลุ่มซึ่งไม่มีความเกี่ยวข้องกัน (Mutually Exclusive) ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูล สำหรับการทดลองนี้จึงได้วัดและประเมินผลการทำงานดังนี้

3.3.1 ทฤษฎีแผนผังการจัดระเบียบตัวเอง

ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [Akinduko, A.A 2016] [Goodfellow, Ian 2016] เป็นอัลกอริทึมนิเวศวิทยาที่นิยมใช้มากที่สุด โดยแนวคิดของ SOM คือ การทำซ้ำข้อมูลเพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ เป็นวิธีเกี่ยวกับการจัดกลุ่มด้วยตัวเองโดยใช้โครงสร้างตาข่ายระบบประสาท เป็นการเรียนรู้แบบไม่มีผู้สอน มีลักษณะการเรียนรู้แบบเครือข่าย 2 ชั้น เป็นการจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติคือการจัดกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้เคียงกัน โดยโครงสร้างการทำงานของ SOM ชั้นแรกทำหน้าที่นำเข้าข้อมูลและจัดส่งข้อมูลให้แก่นิเวศวิทยาชั้นที่สองทุกโหนดในตาราง หรือ อะเรย์นิเวศวิทยา (array of neurons) ระหว่างชั้นจะเชื่อมต่อกันด้วยค่าน้ำหนัก (weight) จากนั้นข้อมูลจะถูกส่งไปยังนิเวศวิทยาชั้นที่สอง เพื่อทำการเปรียบเทียบ

ว่าใกล้เคียงกับค่ากลางกลุ่มใดมากที่สุด แต่ละโหนดในชั้นนี้จะมีความสัมพันธ์กันแบบ เพื่อนบ้าน (neighborhood relation) ทำให้เกิดเป็นรูปแบบ 2 มิติ จากนั้นจะทำการปรับค่า น้ำหนักของตัวที่เป็นผู้ชนะ (winner) ทำให้ข้อมูลที่มีข้อมูลเข้า เข้ามาเกิดการปรับเปลี่ยน และเมื่อผ่านการเรียนรู้ไปหลายๆรอบจะทำให้ ได้



กลุ่มข้อมูลออกมาเป็นผลลัพธ์ จากนั้นจึงนำผลลัพธ์ที่ได้นั้นมาผ่านกระบวนการ Visualization เพื่อแสดงผลลัพธ์ที่ได้นั้นออกมาเป็นกราฟชนิดต่างๆต่อไป

ภาพที่ 3.10 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง

ขั้นตอนการทำงานของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

ทำการจัดกลุ่มของข้อมูลนำเข้าให้อยู่ในกลุ่มข้อมูลที่ใกล้เคียง ทำการ Normalize ค่าข้อมูลแต่ละตัว

เพื่อให้ได้ค่ากลางเป็น
$$x_i = \frac{x_i - \min}{\max - \min}$$
 กำหนดให้ x_1, x_2, \dots, x_n เป็นข้อมูลที่ถูกรับเข้า n คือจำนวน ตัวอย่างข้อมูลเข้า \min คือ ค่าที่น้อยที่สุดของชุดข้อมูล \max คือ ค่าที่มากที่สุดของชุดข้อมูล

กำหนดโครงสร้างโครงข่ายประสาทเทียมในลักษณะสองมิติ กำหนด แนวแกน x และแกน y กำหนดชุดของ ข้อมูล สุ่มค่าน้ำหนักเริ่มต้น ในโครงข่ายดังแสดงในภาพที่ 3.10

สุ่มค่าเริ่มต้นให้กับค่ากลางของกลุ่ม (cluster center) $W = \{w_1, w_2, \dots, w_p\}$, โดยที่ p คือจำนวนกลุ่ม

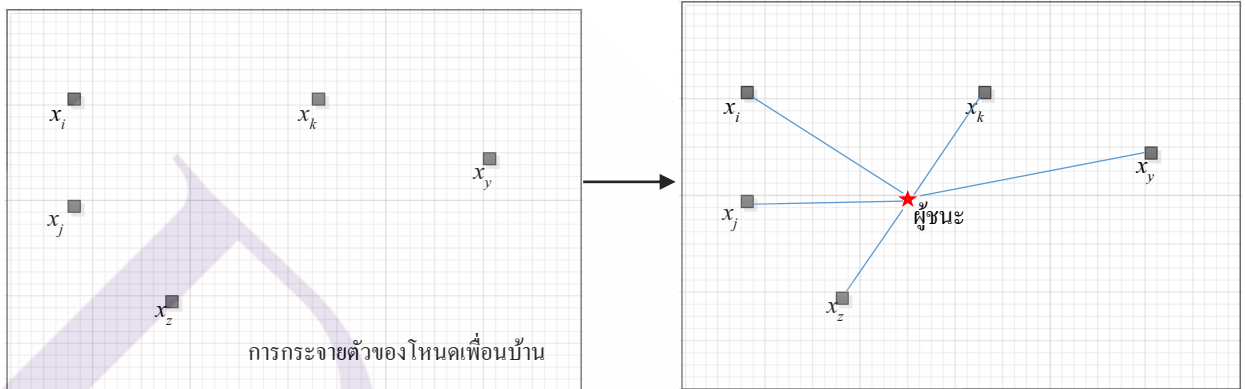
เวกเตอร์ค่ากลางคือ $W_j = \{w_{j1}, w_{j2}, \dots, w_{jm}\}$, เมื่อ $1 \leq j \leq p$ คำนวณหาระยะทางระหว่างโครงข่ายโดย ยูคลิเดียน (Euclidean distance) เพื่อหาผู้ชนะ (winner) ซึ่งจะหาได้จากโครงข่ายที่ใช้ระยะทางที่ได้จากการ

คำนวณ เพื่อหาค่าน้อยที่สุด $C(k_1, k_2) = \min_{i,j} C_{i,j}$ เมื่อ k_1, k_2 คือดัชนีของโหนดผู้ชนะ

$$C_{i,j}^2 = \|x - w_j\|^2 \tag{7}$$

$$C_{i,j}^2 = \sum_{i=1}^w (x_i - w_{ij})^2 \tag{8}$$

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



ภาพที่ 3.11 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน

คำนวณค่ากลางสำหรับกลุ่มที่เป็น ค่าผู้ชนะใหม่

$$w_j(t+1) = w(t) + \eta(x_i - w_j(t)) \tag{9}$$

ทำซ้ำขั้นตอนที่ 4-5 จนครบชุดข้อมูลนำเข้า หรือจนกว่าค่าน้ำหนักเริ่มจะคงที่

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

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

หาระยะห่างระหว่างโหนดนั้นๆ กับ โหนดที่เป็นผู้ชนะ $\rho = \sqrt{(k_1 - i)^2 + (k_2 - j)^2}$

หาโหนดเพื่อนบ้านที่ใกล้เคียงที่สุด โดยขอบเขตจะลดลงตามเวลาตามรูปแบบดังนี้

$$h(\rho, t) = \exp\left(-\frac{\rho^2}{\sigma^2(t)}\right) \left(1 - \frac{2}{\sigma^2(t)} \rho^2\right) \tag{10}$$

$$h(\rho) = \begin{cases} 1, & |\rho| \leq a, \\ -\frac{1}{3}, & a < |\rho| \leq 3a, \\ 0, & |\rho| > 3a, \end{cases} \tag{11}$$

ปรับค่าน้ำหนัก ของแต่ละโหนดด้วย

$$W_{i,j}(t+1) = W_{i,j}(t) + a(t)h(\rho,t)(X^l(t) - W(t)) \quad (12)$$

3.3.2 โครงข่ายประสาทเทียมแบบสังวัตนาการ

โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Networks: CNN) [J. Long 2015] เป็นเพอร์เซ็ปตรอนหลายชั้น (Multi-layer Perceptron) ที่มีโครงสร้างแบบลำดับชั้น (Hierarchical Architectures) ถูกออกแบบมาสำหรับการเรียนรู้ข้อมูลแบบก้าวหน้าในระดับสูงเพื่อใช้ในการจำแนกข้อมูลให้ได้ผลลัพธ์ในขั้นสุดท้าย สามารถเขียนการดำเนินงานเป็นลำดับชั้นพื้นฐานของ CNN ดังนี้

Convolution Layer ลำดับชั้นจะทำการกรองข้อมูลแบบ 2 มิติ ระหว่าง ข้อมูลภาพ I , ตัวกรอง w และ h เป็นข้อมูลภาพที่ถูกสร้างขึ้น ให้ CT แทนความสัมพันธ์ของข้อมูลเข้าและออก เมื่อตัวกรองตอบสนองจากข้อมูลเข้าที่เชื่อมต่อกับข้อมูลออกตัวเดียวกันจะถูกเชื่อมต่อกัน สามารถเขียนเป็นสมการ

$$h_j = \sum_{i,k \in CT_{i,k,j}} (I_i * W_k) \quad (13)$$

Sub-sampling Layer เป็นลำดับชั้นที่ถูกสร้างขึ้นเพื่อดำเนินการ “Max-Pooling” เป็นการเปลี่ยนแปลงและลดทอนให้สามารถทำงานได้อย่างรวดเร็วขึ้น

Fully-connected Layer เป็นลำดับชั้นที่รับข้อมูลเข้าสำหรับทุกลำดับชั้นเพื่อทำการรวมเข้าเป็นข้อมูลออก แบบ 1 มิติสำหรับลำดับชั้นถัดไป

Output Layer เป็นลำดับชั้นข้อมูลออกที่ถูกจำแนกเพียงคลาสเดียวแบบ 1 มิติจากลำดับชั้น Fully-connected Layer

โดยปกติทั่วไปขั้นตอนการเรียนรู้ของ CNN จะใช้วิธีการ Gradient Descent Approaches เพื่อลดความผิดพลาดของทุกลำดับชั้นใน CNN ที่มีการเปลี่ยนแปลงตลอดเวลา แต่บางงานวิจัยจะมีการใช้ขั้นตอน Back-propagation สำหรับการเรียนรู้ด้วย Max-Pooling Convolutional Neural Network และใช้ค่าความคลาดเคลื่อนยกกำลังสอง (Sum of the Squared Error) เพื่อลดความผิดพลาด แต่อย่างไรก็ตามเนื่องจากค่าของข้อมูลโดยทั่วไปเป็นแบบไม่เป็นเชิงเส้น (Non-linear) และมีความผิดพลาดมากวิธีการ Stochastic Gradient Descent จึงเป็นวิธีการที่เหมาะสมในการนำมาใช้สำหรับการทดลองเพื่อหลีกเลี่ยงปัญหาที่เกิดการติดขัดเมื่ออยู่ในจุดที่มีค่าต่ำสุด

3.3.3 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) [Joachims, T. 1998] เป็นขั้นตอนวิธีการที่สามารถนำมาช่วยวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยใช้หลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้นมา เพื่อแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการ

สอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลที่ดีที่สุด กำหนดให้ชุดข้อมูลการเรียนรู้เป็น ข้อมูลภาพ I ที่มีข้อมูลเป็น $\{y_i, x_i\}_{i=1}^m$, เมื่อ $x_i \in R^n$ เป็นรูปแบบข้อมูลเข้าที่ i และ $y_i \in R^n$ เป็น รูปแบบข้อมูลออกที่ i และมีรูปแบบของการจำแนกด้วยวิธีการ SVM ดังสมการ $y(x) = \text{sign}(\sum_{i=1}^m \alpha_i y_i K(x_i, x) + b)$ เมื่อ α_i เป็นค่าบวกคงที่และ b เป็นค่าคงที่จำนวนจริง โดยปกติแล้ว SVM ถูกนำมาใช้กับข้อมูลที่เป็นเชิงเส้น แต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้ในระบบการสอน ให้ระบบเรียนรู้ส่วนใหญ่มักเป็นข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งสามารถแก้ปัญหาดังกล่าวด้วยการ Kernel Function มาใช้จำแนกข้อมูลบนระนาบหลายมิติ และ $K(\dots)$ แทนด้วยค่าข้อมูลแบบไม่เป็นเชิงเส้น Kernel function ที่ถูกปรับค่าเป็นข้อมูลเข้าที่มีหลายมิติ สามารถเขียนสมการของการลดทอน Cost function ได้ดังนี้

$$\text{minimize}_{w, \xi, b} \{ \|w\|^2 / 2 + C \sum_{i=1}^m \xi_i \} \quad (14)$$

ภายใต้ข้อจำกัดของ $y_i(w \cdot x_i - b) \geq 1 - \xi_i$ และ $\xi_i \geq 0$, สำหรับ $i = \{1, 2, \dots, m\}$ เมื่อ ξ_i เป็นข้อมูลใช้วัดตึกการจำแนกข้อมูลที่ผิดพลาดของ x_i และตัวแปรควบคุม C จะควบคุมระหว่างค่า ผิดพลาดและค่าสูงสุดของข้อมูลการเรียนรู้ โดยที่ถ้า $C = \infty$ จะนำไปไปยัง Margin ที่ยากของ SVM โดยทั่วไป SVM จะถูกเรียนรู้ด้วยการคูณของค่าลากรองจ์ (Lagrangian) และการแก้ปัญหาด้วยการเขียน โปรแกรมกำลังสอง (Quadratic Programming) แต่อย่างไรก็ตามยังไม่สามารถใช้ได้กับ CNN ทำให้ต้องมีการ เรียนรู้ CNN และ SVM พร้อมกันโดยใช้เพียงขั้นตอนเดียวโดยการปรับค่าที่เพิ่มขึ้นแบบออนไลน์ด้วยวิธีการที่ เรียกว่า Stochastic Gradient Descent (SGD) [Y. Tsuruoka 2009]

3.3.4 โครงข่ายประสาทเทียมแบบสังวัตนาการ-ซัพพอร์ตเวกเตอร์แมชชีน

โครงข่ายประสาทเทียมแบบสังวัตนาการ-ซัพพอร์ตเวกเตอร์แมชชีน (Convolutional Neural Network - Support Vector Machines: CNNSVM) [J. Long 2015][Ren, S 2015] เป็นการนำสองวิธีการ มาผสมรวมกันโดยที่ CNN จะถูกเรียนรู้ข้อมูลเข้าที่ไม่สัมพันธ์กับรูปแบบของข้อมูล และ SVM ใช้ Kernel Function สามารถจำแนกผลลัพธ์ได้ตามกลุ่มที่ถูกเรียนรู้ได้ ซึ่ง CNN จะสามารถทำงานได้อย่างมีประสิทธิภาพ มากสำหรับข้อมูลภาพคงที่แต่ไม่สามารถจำแนกผลลัพธ์ได้เสมอไป เช่นเดียวกันกับวิธี SVM จะถูกจำกัดด้วยค่า Kernel function ที่ซับซ้อนแต่ผลลัพธ์ของการจำแนกข้อมูลที่ดีด้วยวิธีการ Soft-margin Approaches ดังนั้นสามารถเขียนวิธีการจำแนกได้ดังนี้

กำหนดให้ $f(x)$ เป็นฟังก์ชันการตัดสินใจและสมการทั่วไปของ SVM สามารถเขียนได้เป็น $f(x) = (w \cdot \phi(x) + b)$, เมื่อ w เป็นเวกเตอร์ของค่าน้ำหนัก, b เป็นค่าอคติ (Bias) และทุก พารามิเตอร์จะเพิ่มค่า ϕ แทน Arbitrary Function ดังนั้นวัตถุประสงค์หลักของการจำแนกด้วย CNNSVM เพื่อทำการหาค่าเส้นแบ่งที่เหมาะสมที่สุดสำหรับข้อมูล $f(x)$ ดังนั้นกำหนดให้ชุดข้อมูลการเรียนรู้เป็น

$S = \{(x_i, y_i)\}_{i=1}^m$ เมื่อ $x_i \in R^n$ และ $y_i \in \{+1, -1\}$ สำหรับการจำแนกสองกลุ่ม และค่า Cost function มีสมการเป็น

$$\text{minimize}_w \{ \lambda/2 \|w\|^2 + 1/m \sum_{i=1}^m l(x_i, y_i; w) \}, \quad (15)$$

เมื่อ $\lambda \geq 0$, และ $l(x_i, y_i; w) = \max\{0, 1 - y_i f(x_i)\}$. สำหรับเทอม $\max\{0, 1 - y_i f(x_i)\}$ จะถูกแทนด้วย Hinge-loss และกำหนดให้ $f(x_i)$ มีความสัมพันธ์กับ Kernel $K(\dots)$, และกำหนดให้ $\lambda/2 \|w\|^2$ ถูกใช้เป็นค่าสูงสุดของ Margin เมื่อเทอม $1/m \sum_{i=1}^m l(x_i, y_i; w)$, เป็นค่าต่ำสุดของค่าความคลาดเคลื่อนในการเรียนรู้และให้ ค่า C เป็นพารามิเตอร์ของการจำแนกคลาดเคลื่อน โดยที่มีความสัมพันธ์กับพารามิเตอร์ $\lambda = 1/mC$. โดยทั่วไปแล้วการเรียนรู้ในชุดของข้อมูลของ CNN และ SVM จะต้องทำให้ไปถึงค่าต่ำที่สุด $R(f(x_i), y_i)$ ดังนั้นในสมการที่ (9) จะถูกทำให้มีค่าต่ำสุดด้วยวิธีการ SGD [23] โดยแสดงขั้นตอนด้านล่าง

Algorithm 1: Stochastic Sub-Gradient Projection

Initialize: w such that $\|w\| \leq (1/\sqrt{\lambda})$

For $t = \{1, 2, \dots, T\}$ do

Swarm acquires m samples: $\{A_t(i) | i = 1, 2, \dots, m\}$,

$A_t^+ = \{(X, y) \in A_t : 1 - y(w_t \cdot X) < 1\}$ and w_{t+1} .

$$\eta_t = 1/\sqrt{\lambda} \left(w_{t+\frac{1}{2}} \right) = (1 - \eta_t \lambda) w_t + \frac{\eta_t}{m} \sum_{(x,y) \in A_t^+} yX,$$

$$w_{t+1} = \text{minimize} \left\{ 1, \frac{1/\sqrt{\lambda}}{\|w_{t+\frac{1}{2}}\|} \right\} w_{t+\frac{1}{2}}$$

End

Return $w_{t+\frac{1}{2}}$

ภาพที่ 3.12 Algorithm Stochastic Sub-Gradient Projection

จากภาพที่ 3.12 Algorithm 1 จะมีข้อมูลเข้าพารามิเตอร์ m จากสมการ (15) และค่า T แทนจำนวนครั้งในการวนซ้ำ $t = \{1, 2, \dots, T\}$ ในแต่ละครั้งของการวนซ้ำจะกำหนดค่าน้ำหนักเป็นเวกเตอร์ w ที่มีค่าบรรทัดฐานเป็น $\|w\| \leq (1/\sqrt{\lambda})$ ในแต่ละครั้งของการวนซ้ำ t จะให้เซตของ $\{A_t(i) | i = 1, 2, \dots, m\}$ ต้องผ่านกระบวนการ Swarm ในสมการที่ (15) ก่อนและสมการ

$$f(w; A_t) = \frac{\lambda}{2} \|w\|^2 + \frac{1}{m} \sum_{(x,y) \in A_t} l(x_i, y_i; w) \quad (16)$$

3.3.5 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญเพราะการนำวิธีการที่นำเสนอไป

ข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้
 สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่ม
 กระบวนการแรกใหม่ ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูก
 พิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูล ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของ
 การจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย

1. ค่าความแม่นยำ (False positive rate / Precision: Pr.) เป็นอัตราส่วนของการค้นพบภาพที่
 ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหามาได้

$$Pr = \varepsilon / \partial \quad (17)$$

เมื่อกำหนดให้ ε แทนจำนวนข้อมูลที่เกี่ยวข้องและถูกดึงออกมาได้อย่างถูกต้อง และ ∂ แทนจำนวนข้อมูลที่ถูกดึง
 ออกมาทั้งหมด

2. ค่าความระลึก (True positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจาก
 จำนวนภาพที่ถูกต้องทั้งหมด

$$Re = \varepsilon / \sigma \quad (18)$$

เมื่อ σ แทนจำนวนข้อมูลที่เกี่ยวข้องทั้งหมด

3. ค่าเฉลี่ยของความแม่นยำ (mean average precision: MAP) เป็นค่าเฉลี่ยความแม่นยำของคำหลัก
 ค้นหาคำที่เกี่ยวข้องกัน (relevance) และนำไปจัดอันดับ โดยมีความสัมพันธ์กับคิวรี (query) ค่าเฉลี่ยของ
 ความแม่นยำเมื่อกำหนดค่าสูงสุดของ k ที่มีความเกี่ยวข้องกับข้อมูลภายในที่ถูกค้นคืน สามารถเขียนเป็นกลุ่ม
 ของเซตของข้อมูลที่เกี่ยวข้องได้เป็น $q_i \in Q, \{d_1, d_2, \dots, d_{m_j}\}$ และ R_{jk} เป็นกลุ่มของข้อมูลที่ถูกค้นคืน
 สูงสุดจากข้อมูลภาพ d_k

$$mAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} Pr(R_{jk}) \quad (19)$$

บทที่ 4

ผลการวิจัย

การทดลองด้วยการวัดประสิทธิภาพของการทดลองการจำแนกข้อมูลด้วยความสัมพันธ์ระหว่างวัตถุและแอ็คชัน โดยแบ่งข้อมูลภาพจากฐานข้อมูล Ground Truth และเปรียบเทียบกับข้อมูลภาพบุคคลทั่วไป และภาพผู้สุงวัย โดยที่ภาพที่นำมาทดลองนั้นได้ถูกคัดเลือกมีความสอดคล้องกันของวัตถุและสัมพันธ์เกี่ยวเนื่องกัน ทั้งบุคคลในภาพเพื่อให้เห็นผลอย่างชัดเจนว่ามีประสิทธิภาพเป็นอย่างไร สอดคล้องกับความต้องการเพียงใดเมื่อนำมาใช้งานจริง สำหรับข้อมูลภาพที่ใช้ในการทดลองทั้งหมดจะคัดเลือกภาพที่เห็นวัตถุอย่างเด่นชัดและมีความสอดคล้องกับฐานข้อมูลพื้นฐานจากฐานข้อมูลภาพแอ็คชันจาก PASCAL2 [Everingham, M 2010] Activitynet [F. Caba Heilbron 2015] และ ImageNet [J. Deng, 2009] โดยที่แต่ละกลุ่มคำจะมีแอ็คชันที่มีความสอดคล้องกันทั้งภาพที่อยู่ในกลุ่มที่คาดหวัง และภาพที่ไม่อยู่ในกลุ่มซึ่งไม่มีความเกี่ยวข้องกัน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูล เพื่อทำการจำแนกข้อมูลภาพ ประกอบด้วย (1) ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [Akinduko, A.A 2016] [Goodfellow, Ian 2016] (2) โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Networks: CNN) [J. Long 2015] (3) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) [Joachims, T. 1998] (4) โครงข่ายประสาทเทียมแบบสังวัตนาการ-ซัพพอร์ตเวกเตอร์แมชชีน (Convolutional Neural Network - Support Vector Machines: CNNSVM) [J. Long 2015][Ren, S 2015] (5) Algorithm Stochastic Sub-Gradient Projection สำหรับการทดลองจำแนกภาพเพื่อให้เกิดประสิทธิภาพสูงสุดจะทำการทดลองเป็น 3 ส่วน ดังนี้

4.1 การวัดประสิทธิภาพการจำแนกข้อมูลภาพจากฐานข้อมูลมาตรฐาน Ground Truth

การเปรียบเทียบข้อมูลทั้งหมดจากฐานข้อมูลภาพทั่วไปที่เป็นมาตรฐานด้วย Ground Truth [J. Wu 2015] ด้วยการใช้คำหลักเพียงอย่างเดียว ข้อมูลภาพทั้งหมดถูกคัดเลือกตามกลุ่มคำศัพท์หลักทั้งหมด 4,800 ภาพแบ่งเป็น 16 กลุ่มคำศัพท์ กลุ่มละ 300 ภาพจากฐานข้อมูลภาพแอ็คชัน Pascal2 [Everingham, M.2010] และ Activitynet [F. Caba Heilbron 2015] ข้อมูลภาพที่ใช้ในการทดลองทั้งหมดถูกคัดเลือกเฉพาะภาพที่เห็นวัตถุอย่างเด่นชัด และมีความสอดคล้องกับฐานข้อมูลพื้นฐานที่ประกอบด้วย 16 คำศัพท์หลัก ดังนี้ aero, bike, bird, boat, bus, car, cat, chair, cow, table, dog, horse, plant, sofa, train, และ television การทดลองเริ่มจากการเปรียบเทียบการจำแนกข้อมูลภาพตามกลุ่มภาพ Ground Truth จากฐานข้อมูล Pascal2 สำหรับการทดสอบประสิทธิภาพการจำแนกข้อมูลภาพ โดยข้อมูลกลุ่มคำศัพท์ถูกคัดเลือกและมีพื้นฐานการทำลองเพื่อทำการเปรียบเทียบมาจากการทดลองของ J. Wu [J. Wu 2015] ที่มีการคัดเลือกวัตถุบนภาพจาก 20 กลุ่มและทำเป็นมาตรฐานของคำศัพท์ที่น่าเชื่อถือ ดังนั้นจากตารางที่ 4.1 จึงได้ใช้วัตถุและวิธีเปรียบเทียบพื้นฐานทั้งหมด 6 วิธีดังนี้ Generalized Hierarchical Matching (GHM) [J. Wu 2015][Q. Chen 2012], Ambiguity Guided Graph Shift (AGS) [J. Wu 2015] [J. Don] 2013, NUS [J. Wu 2015] [Z. Song

2011], Convolutional Neural Network (CNN), Support Vector Machines (SVM) และ Convolutional Neural Network - Support Vector Machines (CNNSVM) ดังแสดงผลลัพธ์ในตารางที่ 4.1 แสดงการจำแนกตามกลุ่มคำศัพท์จะเห็นว่าวิธี AGS ได้ค่าเฉลี่ยของความแม่นยำถึง 71.1% ซึ่งมีความใกล้เคียงกับวิธีการ CNNSVM ที่นำเสนอ 71.2% การจำแนกข้อมูลสำหรับการเรียนรู้ด้วย CNN ได้กำหนดอัตราการการเรียนรู้ถูกกำหนดไว้ที่ 0.001 สำหรับ CNN และ 0.002 สำหรับ CNNSVM

ตารางที่ 4.1 ผลการจำแนกข้อมูลภาพกับวิธีมาตรฐานด้วยข้อมูล Ground Truth *[J. Wu 2015]

Method	Objects Ground Truth																			mAP (%)	
	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train		tv
GHM*	76.7	74.7	53.8	72.1	40.4	71.7	83.6	66.5	52.5	57.5	62.8	51.1	81.4	71.5	86.5	36.4	55.3	60.6	80.6	57.8	64.7
AGS*	82.2	83.0	58.4	76.1	56.4	77.5	88.8	69.1	62.2	61.8	64.2	51.3	85.4	80.2	91.1	48.1	61.7	67.7	86.3	70.9	71.1
NUS*	82.5	79.6	64.8	73.4	54.2	75.0	77.5	79.2	46.2	62.7	41.4	74.6	85.0	76.8	91.1	53.9	61.0	67.5	83.6	70.6	70.5
CNN	73.5	69.3	50.8	70.3	34.7	66.7	71.8	62.5	42.2	44.7	37.0	50.0	57.9	63.9	71.1	46.4	52.3	43.4	73.6	58.8	57.0
SVM	81.4	70.3	76.1	71.3	40.7	62.3	69.8	68.1	50.7	49.7	40.7	55.0	52.3	65.2	76.1	42.5	59.0	48.1	87.2	71.8	61.9
CNNSVM	86.6	79.9	75.3	78.5	48.4	75.8	83.5	78.0	55.1	61.6	55.7	68.0	78.9	75.9	86.5	49.5	67.3	61.6	87.1	70.3	71.2

4.2 การวัดประสิทธิภาพการจำแนกกิจกรรมด้วยความสัมพันธ์ระหว่างวัตถุและแอ็คชัน

การทดลองสำหรับเปรียบเทียบด้วยความสัมพันธ์จากการทำนาย ด้วยการวัดประสิทธิภาพของการทดลองการจำแนกกิจกรรมด้วยความสัมพันธ์ระหว่างวัตถุและแอ็คชัน โดยใช้ข้อมูลภาพจาก 4.1 และทำการแบ่งข้อมูลภาพจากฐานข้อมูล Ground Truth ข้างต้นเป็น 2 ชุด ชุดละ 2,800 ภาพ โดยที่ข้อมูลภาพใน Set I และ Set II จะต้องมีความสัมพันธ์ของวัตถุภายในภาพที่ประกอบด้วย แบบ implied-by, แบบ type-of และแบบ mutually exclusive ดังแสดงผลการทดลองดังแสดงในตารางที่ 4.2 จะเห็นว่า การทดลองด้วยชุดข้อมูล Set I ประกอบด้วย วิธี Softmax [Goodfellow, Ian 2016] [Akinduko, A.A 2016] ที่มีการกำหนดความสัมพันธ์ของวัตถุอยู่ที่ 30 กิจกรรม โดยได้คัดเลือกกิจกรรมที่สำคัญจาก Activitynet [F. Caba Heilbron 2015] โดยแบ่งเป็นกลุ่มกิจกรรมใหญ่ๆได้ดังนี้ Food/kitchen, Playing sport, Playing musical, Relaxing and Leisure, Housework ดังแสดงในภาพที่ 4.1 เพื่อใช้ในการทดลองเบื้องต้นโดยแบ่งภาพกิจกรรมสำหรับการทดลองเป็น 2 ชุด การจำแนกข้อมูลสำหรับการเรียนรู้ด้วย CNN ได้กำหนดอัตราการการเรียนรู้ถูกกำหนดไว้ที่ 0.001 สำหรับ CNN และ 0.002 สำหรับ CNNSVM

ผลการทดลองสำหรับฐานข้อมูลชุด Set I สามารถจำแนกกิจกรรมต่างๆด้วยการใช้วัตถุและความสัมพันธ์ของแอ็คชันที่เกิดขึ้น ได้ค่าเฉลี่ย mAP ของการจำแนกด้วย Softmax อยู่ที่ 33.2% และ CNN, SVM, CNNSVM อยู่ที่ 43.2%, 51.4% และ 64.4% ตามลำดับ ส่วนวิธีการที่นำเสนอด้วยการจำแนกแบบ CNNSVM ที่มีการใช้ SGP ของชุดทดลอง Set I จะได้ค่าเฉลี่ย mAP อยู่ที่ 68.3% จะเห็นว่ามีค่าเฉลี่ยของความแม่นยำที่เพิ่มขึ้นเกือบ 4% แต่สำหรับการจำแนกด้วยวิธีการ CNNSVM-SGP กิจกรรม Doing karate กับกิจกรรม Doing gymnastics จะมีค่าเฉลี่ยของความแม่นยำที่ค่อนข้างต่ำเพียง 65.9% และ 65.7% เท่านั้น เนื่องจากความสัมพันธ์ระหว่างวัตถุที่เกิดขึ้นภายในภาพอาจจะมีน้อยหรือไม่เด่นชัดเท่าที่ควรจึงทำให้มีค่าความถูกต้องที่ต่ำกว่ากิจกรรมอื่น



Eating and Drinking



Food preparation



Washing dishes

ก. Food/kitchen



Doing karate



Playing volleyball



Playing hockey

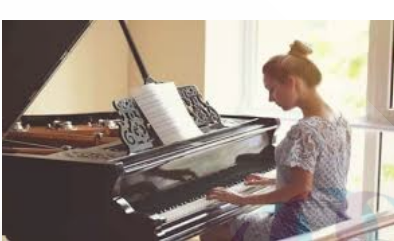
ข. Playing Sport



Playing saxophone



Playing violin



Playing piano

ค. Playing musical



Walking the sea



Painting



Smoking cigarette

ง. Relaxing and Leisure

ภาพที่ 4.1 ตัวอย่างภาพในกลุ่มกิจกรรม

จากตารางที่ 4.3 ชุดข้อมูล Set II จะเห็นว่าเมื่อใช้วิธีการจำแนกด้วย CNNSVM-SGP ค่าเฉลี่ยของความแม่นยำที่เพิ่มขึ้นถึง 77.3% เมื่อเปรียบเทียบการจำแนกด้วยวิธีการของ CNNSVM ได้ค่าเฉลี่ยของความแม่นยำ 69.5% จะแสดงให้เห็นว่าวิธีการ SGP สามารถเพิ่มประสิทธิภาพของการจำแนกได้สูงขึ้นเกือบ 8% สำหรับชุดข้อมูล Set II จะเห็นว่าการสร้างความสัมพันธ์ของระหว่างวัตถุและแอ็คชั่นที่เกิดขึ้นบนภาพ มาสร้าง

ความสัมพันธ์ภายใน และเป็นการทำนายความสัมพันธ์จากการเรียนรู้รูปแบบของแอ็คชั่นที่เกิดขึ้น และการเก็บค่าน้ำหนักที่ได้จากความสัมพันธ์ระหว่างกันสามารถทำให้การใช้ SGP ของชุดทดลอง Set I และ Set II มีค่าเฉลี่ยของความแม่นยำที่เพิ่มขึ้น

ตารางที่ 4.2 ผลการจำแนกกิจกรรมด้วยความสัมพันธ์ของวัตถุและแอ็คชั่นด้วยชุดทดลอง Set I

Activities		Classification Method					
		Softmax	CNN	SVM	CNNSVM	CNNSVM-SGP	Average mAP(%)
Food/kitchen	Eating and Drinking	31.7	40.5	52.3	62.5	74.8	52.3
	Food preparation	34.6	42.0	53.9	62.5	70.7	52.7
	Washing dishes	33.0	40.7	52.6	70.9	66.5	52.7
Playing Sport	Doing aerobics	33.8	41.9	46.3	69.4	68.5	52.0
	Doing karate	31.1	45.0	47.8	63.8	65.9	50.7
	Ta chi	30.2	36.0	50.7	67.6	71.6	51.2
	High jump	31.1	48.8	49.8	69.2	69.6	53.7
	Playing volleyball	31.6	48.6	53.4	61.1	68.4	52.6
	Playing kickball	31.5	46.1	50.0	67.6	68.0	52.6
	Weightlifting	39.0	40.1	52.7	63.5	71.6	53.4
	Playing hockey	36.9	46.0	47.7	63.3	66.9	52.2
	Doing gymnastics	32.8	49.5	55.8	63.9	65.7	53.5
	Playing squash	33.9	43.9	52.6	64.7	74.3	53.9
	Playing badminton	30.2	45.7	48.4	64.8	68.5	51.5
	Playing polo	30.7	46.9	55.9	64.3	69.0	53.4
	Horseback riding	32.9	46.6	52.5	61.7	69.2	52.6
Playing musical	Playing saxophone	32.0	41.1	51.3	63.8	66.7	51.0
	Playing guitar	32.8	44.0	51.6	65.7	69.8	52.8
	Playing piano	31.7	43.7	55.9	62.3	65.9	51.9
	Playing violin	37.1	43.0	51.5	62.9	66.4	52.2
Relaxing and Leisure	Walking sea	33.1	39.0	50.3	65.0	65.3	50.5
	Lying down	32.9	40.1	50.9	63.9	65.3	50.6
	Walking dog	33.5	48.2	51.6	64.0	65.5	52.6
	Painting	39.2	42.0	46.9	65.1	66.4	51.9
	Smoking cigarette	30.8	44.8	54.6	67.1	65.8	52.6
Housework	Cleaning window	35.7	38.0	47.7	61.2	65.6	49.6
	Laundry	35.7	42.0	51.6	61.0	69.5	51.9
	Lawn garden	30.1	36.0	52.0	61.2	68.5	49.6
	Care for pets	33.8	43.7	51.8	63.9	70.2	52.7
	Fixing bicycle	31.8	43.5	50.9	63.2	69.5	51.8
Average mAP(%)		33.2	43.2	51.4	64.4	68.3	52.1

ตารางที่ 4.3 ผลการจำแนกกิจกรรมด้วยความสัมพันธ์ของวัตถุและแอ็คชั่น ชุดทดลอง Set II

Activities		Classification Method					Average mAP(%)
		Softmax	CNN	SVM	CNNSVM	CNNSVM-SGP	
Food/kitchen	Eating and Drinking	34.3	35.9	43.9	68.5	73.0	51.1
	Food preparation	35.2	41.6	43.7	73.5	75.0	53.8
	Washing dishes	37.3	41.2	46.8	69.5	76.2	54.2
Playing Sport	Doing aerobics	32.4	42.3	46.3	72.2	75.7	53.8
	Doing karate	29.5	40.0	52.0	71.2	72.6	53.0
	Ta chi	31.1	37.9	43.9	73.1	71.8	51.6
	High jump	34.8	37.9	45.3	71.5	77.3	53.4
	Playing volleyball	35.6	36.9	47.0	64.7	78.6	52.6
	Playing kickball	32.8	42.9	51.1	72.2	79.3	55.7
	Weightlifting	33.6	40.9	44.6	72.9	74.0	53.2
	Playing hockey	38.0	40.3	43.1	71.0	80.1	54.5
	Doing gymnastics	33.5	42.0	51.8	66.6	78.5	54.5
	Playing squash	31.0	38.0	51.9	71.2	77.5	53.9
	Playing badminton	29.7	41.5	44.0	64.1	78.3	51.5
	Playing polo	28.0	38.3	42.4	66.0	79.3	50.8
	Horseback riding	33.9	34.2	43.0	67.0	83.0	52.2
	Playing musical	Playing saxophone	31.9	40.3	47.9	69.0	78.2
Playing guitar		28.8	41.6	50.6	70.8	81.4	54.7
Playing piano		36.6	37.4	49.1	67.8	76.3	53.4
Playing violin		28.3	40.1	43.7	68.3	78.3	51.7
Relaxing and Leisure	Walking sea	36.9	40.9	51.0	69.0	76.3	54.8
	Lying down	31.4	38.8	47.8	68.0	77.6	52.7
	Walking dog	31.9	35.9	47.9	64.9	76.4	51.4
	Painting	30.8	39.1	42.8	66.0	77.4	51.2
	Smoking cigarette	31.1	38.2	48.3	69.0	78.4	53.0
Housework	Cleaning window	34.7	37.0	49.8	73.8	79.0	54.9
	Laundry	32.0	37.9	49.9	69.4	77.9	53.4
	Lawn garden	31.7	36.6	48.8	69.5	71.9	51.7
	Care for pets	33.4	42.0	45.0	70.9	79.1	54.1
	Fixing bicycle	34.2	38.4	43.9	72.6	80.1	53.8
Average mAP(%)		32.8	39.2	46.9	69.5	77.3	53.1

จากตารางที่ 4.3 แสดงการจำแนกกิจกรรมต่างๆถึง 30 กิจกรรม ในชุดทดลองที่ 2 จะเห็นว่าวิธีการจำแนกด้วย CNNSVM-SGP สำหรับกิจกรรม Horseback riding มีค่าความแม่นยำที่สูงที่สุดถึง 83% รองลงมาคือ กิจกรรม Playing guitar 81.4% สำหรับกิจกรรม Fixing bicycle และ Playing hockey มีค่าความแม่นยำถึง 80.1% และกิจกรรม Playing polo มีค่าความแม่นยำ 79.3% และกิจกรรม Playing badminton มีค่าความแม่นยำ 78.3% จากที่กล่าวมาสังเกตถึงค่าความแม่นยำที่ได้รับจากการใช้วิธีการจำแนกด้วย CNNSVM-

SGP เป็นการวัดความสัมพันธ์ของกราฟและค่าน้ำหนักของความสัมพันธ์ภายในของวัตถุอย่างเด่นชัด และกิจกรรมที่มีความสัมพันธ์กับวัตถุภายในภาพน้อยจะมีค่าความแม่นยำที่น้อยตาม เช่น กิจกรรม Doing karate และกิจกรรม Ta chi ได้เพียง 72.6% และ 71.8% ตามลำดับ เช่นเดียวกับตารางที่ 4.2

ตารางที่ 4.4 ผลการจำแนกข้อมูลภาพด้วยข้อมูลผู้สูงอายุ ชุดทดลอง Set I

Method	mAP(%) Set I		Average mAP(%)
	ผู้สูงอายุได้รับการดูแล	ผู้สูงอายุที่ควรเฝ้าระวัง	
SVM	58.09	57.76	57.92
CNN	57.24	61.34	59.29
CNNSVM	65.98	68.76	67.37
CNNSVM-SSGP	72.87	78.52	75.69

ตารางที่ 4.5 ผลการจำแนกข้อมูลภาพด้วยข้อมูลผู้สูงอายุ ชุดทดลอง Set II

Method	mAP(%) Set II		Average mAP(%)
	ผู้สูงอายุได้รับการดูแล	ผู้สูงอายุที่ควรเฝ้าระวัง	
SVM	49.37	51.23	50.30
CNN	52.53	54.27	53.4
CNNSVM	67.82	65.03	66.42
CNNSVM-SSGP	74.23	73.98	74.10

4.3 การวัดประสิทธิภาพจากการนำไปประยุกต์ใช้กับการเฝ้าระวังและช่วยเหลือผู้สูงอายุ

สำหรับข้อมูลภาพที่ใช้ในการทดลองทั้งหมด 400 ภาพ คัดเลือกฐานข้อมูลภาพกิจกรรมเฉพาะภาพผู้สูงอายุจาก Activitynet [F. Caba Heilbron 2015] และ ImageNet [J. Deng, 2009] โดยจัดให้อยู่ในหมวดหมู่และทำการแท็กข้อมูลภาพสำหรับการทดลองเพื่อให้อยู่ในกลุ่มและคำหลักที่สามารถใช้ทดลองได้ ภาพที่มีการโพกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน ถ้าข้อมูลภาพไม่สอดคล้องกันจะไม่ถูกนำมาพิจารณา ข้อมูลภาพทั้งหมดถูกคัดเลือกและสุ่มแบ่งกลุ่มเป็น 2 ชุดข้อมูล ชุดละ 200 ภาพ สำหรับการทดลองในแต่ละชุดข้อมูล โดยทำการทดลองจำแนกด้วย วิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) โครงข่ายประสาทเทียมแบบสังวัตนาการ (CNN) และโครงข่ายประสาทเทียมแบบสังวัตนาการ-ซัพพอร์ตเวกเตอร์แมชชีน (CNNSVM) และ CNNSVM ด้วย Stochastic Sub-Gradient Projection (SSGP) ทำการจำแนกออกเป็น 2 กลุ่มคือ ภาพของผู้สูงอายุได้รับการดูแล และภาพของผู้สูงอายุที่ควรเฝ้าระวัง การจำแนกความหมายของกิจกรรมผู้สูงอายุ ได้แสดงผลการทดลองดังแสดงในตารางที่ 4.4 และ 4.5 เป็นการเปรียบเทียบวิธีการจำแนก 4 วิธีการดังนี้ SVM, CNN, CNNSVM และ CNNSVM-SSGP ด้วยการทำนายความสัมพันธ์กิจกรรมจากการปรับค่าน้ำหนัก จะเห็นว่า SVM ในชุดข้อมูลที่ 1 และ 2 จะมีค่า mAP เฉลี่ยอยู่ที่ 57.92%, 50.30% ตามลำดับ การจำแนกด้วย CNN จากผลการทดลองจะเห็นว่ามามีค่าเฉลี่ย mAP ที่สูงกว่า SVM และวิธีการ CNNSVM เมื่อเปรียบเทียบกับข้อมูลทั้ง 2 ชุด มีแนวโน้มไปในทิศทางเดียวกันคือ และ CNNSVM ด้วย SSGP สำหรับชุดข้อมูล Set 1 จะมีค่าเฉลี่ยของ mAP ที่สูงกว่า 67.37%

และ 75.69% ตามลำดับ สำหรับชุดข้อมูล Set 2 ค่า mAP ที่มีการจำแนกด้วย CNNSVM สำหรับผู้สูงอายุได้รับการดูแล 67.82% และ ผู้สูงอายุที่ควรเฝ้าระวัง 65.03% และ การจำแนกด้วย CNNSVM-SSGP อยู่ที่ 74.23% และ 73.98%



บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอรูปแบบการจำแนกความหมายภาพด้วยการแทนข้อมูลภาพด้วยคำศัพท์จากฐานข้อมูล WordNet และกิจกรรมด้วย คลาส Activitynet และแทนความสัมพันธ์ของข้อมูลวัตถุภายในภาพพร้อมกับสร้างการเชื่อมโยงกันภายในภาพเพื่อให้เกิดความหมายการกระทำที่เกิดขึ้น โดยใช้อาศัยการเรียนรู้ความสัมพันธ์ของวัตถุและสภาพแวดล้อมภายในภาพที่เกิดขึ้น เพื่อนำมาประยุกต์ความสัมพันธ์ในการแจ้งเตือนเมื่อเกิดความเสี่ยงของผู้สูงอายุขึ้น

สรุปผลการวิจัย

1. วัตถุประสงค์ของการวิจัย

- 1.1 เพื่อสร้างแบบจำลองวิเคราะห์ ความหมายภาพรูปแบบใหม่
- 1.2 สามารถนำเทคนิคใหม่ที่นำเสนอมาใช้ในการค้นหาความหมายของภาพดิจิทัล
- 1.3 เพื่อเฝ้าระวังสำหรับผู้สูงอายุที่ต้องดูแลเป็นพิเศษ

2. วิธีดำเนินการวิจัย

2.1 การทดลองทั้งหมดจะคัดเลือกภาพที่เห็นวัตถุอย่างเด่นชัดและมีความสอดคล้องกับฐานข้อมูลพื้นฐานจากฐานข้อมูลภาพแอ็คชันจาก PASCAL2 [Everingham, M 2010] Activitynet [F. Caba Heilbron 2015] และ ImageNet [J. Deng, 2009] โดยที่แต่ละกลุ่มคำจะมีแอ็คชันที่มีความสอดคล้องกันทั้งภาพที่อยู่ในกลุ่มที่คาดหวัง และภาพที่ไม่อยู่ในกลุ่มซึ่งไม่มีความเกี่ยวข้องกัน

2.2 เครื่องมือการวิเคราะห์ข้อมูลเพื่อทำการจำแนกข้อมูลภาพ ประกอบด้วย (1) ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) (2) โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Networks: CNN) (3) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) (4) โครงข่ายประสาทเทียมแบบสังวัตนาการ-ซัพพอร์ตเวกเตอร์แมชชีน (Convolutional Neural Network - Support Vector Machines: CNNSVM) (5) Algorithm Stochastic Sub-Gradient Projection (SGP)

2.3 ลำดับการทดลองประสิทธิภาพการจำแนกข้อมูล แบ่งเป็น 3 ขั้นตอนดังนี้

(1) การวัดประสิทธิภาพการจำแนกข้อมูลภาพจากฐานมาตรฐาน Ground Truth ประเมินผลด้วยวิธีการ Generalized Hierarchical Matching (GHM), Ambiguity Guided Graph Shift (AGS), Convolutional Neural Network (CNN), Support Vector Machines (SVM) และ Convolutional Neural Network - Support Vector Machines (CNNSVM)

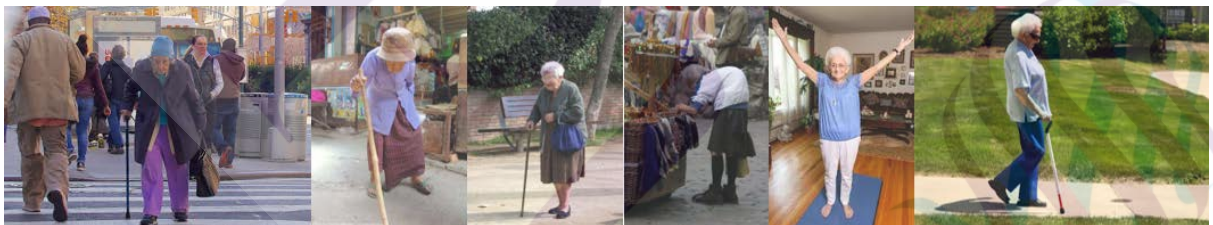
(2) การวัดประสิทธิภาพการจำแนกกิจกรรมด้วยความสัมพันธ์ระหว่างวัตถุและแอ็คชัน ด้วย 5 วิธีการ SOM, CNN, SVM, CNNSVM, และ CNNSVM-SSGP

(3) การวัดประสิทธิภาพจากการนำไปประยุกต์ใช้กับการเฝ้าระวังและช่วยเหลือผู้สูงอายุ ด้วย 4 วิธีการ CNN, SVM, CNNSVM, และ CNNSVM-SSGP

3. ผลการวิเคราะห์ข้อมูล สำหรับการเฝ้าระวังสำหรับผู้สูงอายุที่ต้องดูแลเป็นพิเศษ โดยการใช้วิธีการทำนายความสัมพันธ์แอ็คชัน และการจำแนกภาพผู้สูงอายุด้วยวิธีการ CNNSVM ด้วย SSGP สำหรับชุดข้อมูล Set 1 และ ชุดข้อมูล Set 2 ค่า mAP ประมาณ 75% สำหรับการแจ้งเตือนเมื่อผู้สูงอายุมีความเสี่ยงที่จะเกิดอุบัติเหตุ จะเห็นว่าสามารถนำวิธีการเพื่อไปประยุกต์ใช้ในส่วนต่างๆของระบบการเฝ้าระวังได้



ก.ภาพผู้สูงอายุได้รับการดูแล



ข. ภาพผู้สูงอายุที่ควรเฝ้าระวัง

ภาพที่ 5.1 ตัวอย่างภาพผู้สูงอายุที่ได้รับการดูแล และควรเฝ้าระวัง

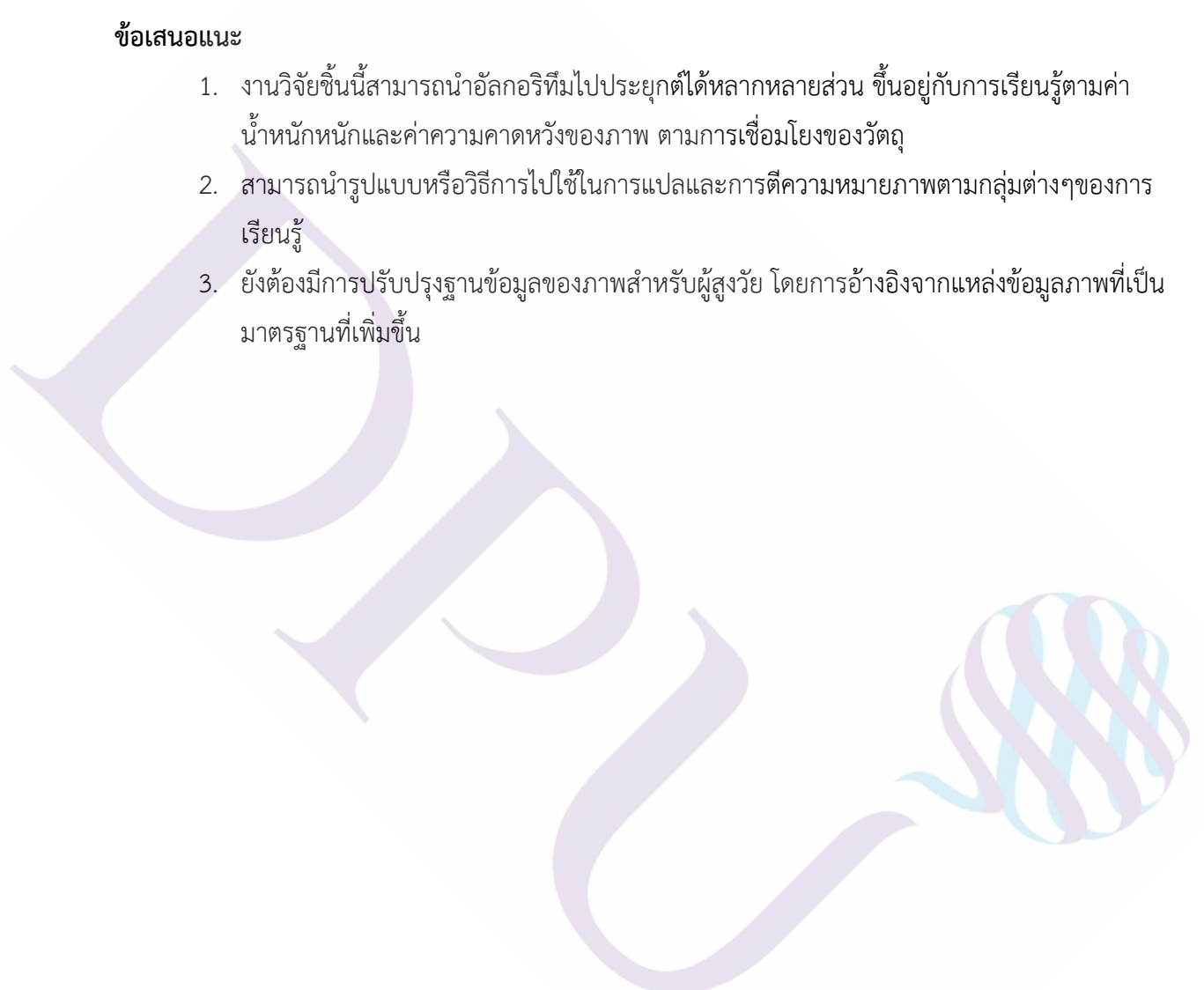
อภิปรายผล

จากผลการทดลองด้วยภาพกิจกรรมของผู้สูงอายุจะเห็นว่าสามารถทำการแจ้งเตือนความเสี่ยงของกิจกรรมผู้สูงอายุได้ ว่ามีการได้รับการดูแลหรือเกิดความเสี่ยงในการทำกิจกรรมประจำวัน เพื่อเป็นแนวทางการป้องกันเบื้องต้นได้ วิธีการนี้ได้พยายามดัดแปลงการทำนายความสัมพันธ์ของการแอ็คชัน ด้วยวิธีการจำแนกแบบ CNNSVM ด้วยอัลกอริทึม SSGP เพื่อทำการจำแนกความสัมพันธ์วัตถุที่ถูกสร้างขึ้นในรูปแบบของค่าน้ำหนักที่ทำให้เกิดการจำแนกที่ดีกว่าการใช้แต่คำศัพท์เพียงอย่างเดียว ซึ่งการทดลองได้มีการแยกออกเป็น 3 ส่วนเริ่มจากการใช้คำศัพท์ทั่วไปโดยมีพื้นฐานตามข้อมูลภาพทั่วไป และมีการปรับและเพิ่มภาพตามกิจกรรมสำคัญของบุคคลตามคลาสที่ถูกกำหนดไว้ โดยมีการประยุกต์กับอัลกอริทึม SSGP และความสัมพันธ์ของการทำนายกิจกรรม และนำอัลกอริทึมมาปรับใช้กับภาพผู้สูงอายุ สามารถทำการจำแนกได้ค่าความแม่นยำสูงถึง 75% ซึ่ง

เบื้องต้นยังสามารถนำอัลกอริทึมไปประยุกต์ใช้เพื่อการเรียนรู้ในเรื่องอื่นที่สามารถสร้างความสัมพันธ์และบ่งบอกเป็นกิจกรรมได้ แต่อย่างไรก็ตามจำนวนกิจกรรมของผู้สูงอายุอาจต้องมีการแยกแยะเพิ่มมากขึ้น รวมทั้งความสัมพันธ์ที่กำหนดขึ้นมาเพื่อเป็นข้อมูลเข้าอาจยังไม่ครอบคลุม ทั้งหมดของฐานข้อมูลภาพ ดังนั้นอาจจำเป็นต้องมีการเพิ่มเติมแอ็คชั่นที่มีความเกี่ยวข้องและความสัมพันธ์ที่จำเป็นเพิ่มขึ้นเพื่อใช้สำหรับการทดลองต่อไป

ข้อเสนอแนะ

1. งานวิจัยชิ้นนี้สามารถนำอัลกอริทึมไปประยุกต์ได้หลากหลายส่วน ขึ้นอยู่กับการเรียนรู้ตามค่าน้ำหนักหนักและค่าความคาดหวังของภาพ ตามการเชื่อมโยงของวัตถุ
2. สามารถนำรูปแบบหรือวิธีการไปใช้ในการแปลและการตีความหมายภาพตามกลุ่มต่างๆของการเรียนรู้
3. ยังต้องมีการปรับปรุงฐานข้อมูลของภาพสำหรับผู้สูงอายุ โดยการอ้างอิงจากแหล่งข้อมูลภาพที่เป็นมาตรฐานที่เพิ่มขึ้น



บรรณานุกรม

- A. F. Bobick and A. D. Wilson, (1997). A state-based approach to the representation and recognition of gesture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12). 1325–1337.
- Akinduko, A.A. Mirkes, E.M. and Gorban, A.N. (2016). SOM: Stochastic initialization versus principal components. *Information Sciences*. 364-365: 213–221.
- Albert Haque, and et al. (2017). Towards vision based smart hospitals: A system for tracking and monitoring hand hygiene compliance. In *Proceedings of Machine Learning for Healthcare JMLR W&C*, 68.
- B. Fosty, C. Crispim-Junior, J. Badie, F. Bremond and M. Thonnat. (2013). Event Recognition System for Older People Monitoring Using an RGB-D Camera. In the 2nd Workshop on Assistance and Service Robotics in a Human Environment in conjunction with IEEE/RSJ International Conference on Intelligent Robots and Systems IROS, ASROB-2013.
- Bangpeng Yao and Li Fei-Fei (2010). Grouplet: A structured image representation for recognizing human and object interactions, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 9–16.
- C. Crispim-Junior, B. Fosty, R. Romdhane, F. Bremond and M. Thonnat. (2013). Combining Multiple Sensors for Event Recognition of Older People. In the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare, MIIRH.
- D. Minnen, I. A. Essa and T. Starner. (2003). Expectation grammars: Leveraging high-level expectations for activity recognition. In *CVPR(2)*, IEEE Computer Society. 626–632.
- Escalante, H. J., Hernández, C., Gonzalez, J., López, A., Montes, M., Morales, E., Sucar, E., , L., Grubinger. (2009). M.: The Segmented and Annotated IAPR TC-12 Benchmark. *Computer Vision and Image Understanding*.
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. (2010). *International Journal of Computer Vision*, 88(2), 303-338.
- F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles. (2015). Activitynet: A large-scale video benchmark for human activity understanding. In *CVPR*.
- F. Negin, M. Koperski, C. Crispim-Junior, F. Bremond, S. Cosar and K. Avgerinakis. (2016). A Hybrid Framework for Online Recognition of Activities of Daily Living In Real-World Settings. In *Proceedings of the 13th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS*.

- George A. Miller (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*.3. 235–312.
- George A. Miller , Claudia Leacock , Randee Teng , Ross T. Bunker (1993). A semantic concordance, *Proceedings of the workshop on Human Language Technology*, March 21-24.
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). "6.2.2.3 Softmax Units for Multinoulli Output Distributions". *Deep Learning*. MIT Press. pp. 180–184.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. (2009). Imagenet: A large-scale hierarchical image database. In *CVPR*.
- J. Dong, W. Xia, Q. Chen, J. Feng, Z. Huang, and S. Yan. (2013). Subcategory-aware object classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- J. F. Allen, and G. Ferguson. (1994). Actions and events in interval temporal logic. *Journal of Logic and Computation*. 4(5), 531–579.
- J. Long, E. Shelhamer, and T. Darrell (2015). Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- J. M.Siskind. (2001). Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *Journal of Artificial Intelligence Research (JAIR)*. 15. 31–90.
- J. Wu, Y. Yu, C. Huang, and K. Yu. (2015). Deep multiple instance learning for image classification and autoannotation, In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Joachims, T. (1998). Text categorization with support vector machines—learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning*, Chemnitz, Germany. (Berlin: Springer),137–142.
- Lillo, I., Soto, A., and Niebles, J. C. (2014). “Discriminative hierarchical modeling of spatio-temporally composable human activities,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH)*, 812–819.
- M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman (2010). The PASCAL Visual Object Classes (VOC) Challenge, *International journal of computer vision*, 88(2), 303–338.
- Michalis Vrigkas, Christophoros Nikou and Ioannis A. Kakadiaris. (2015). A Review of Human Activity Recognition Methods *Front. Robot. AI*.
- N. T. Nguyen, D. Q. Phung, S Venkatesh and H.H. Bui. (2005). Learning and detecting activities from movement trajectories using the hierarchical hiddenMarkov models. In *CVPR(2) IEEE Computer Society*. 955–960.

- Nadia Zouba, François Bremond, Monique Thonnat and Van Thinh Vu. (2007). Multi-sensors Analysis for Everyday Elderly Activity Monitoring. In the SETIT 2007, 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications – Tunis.
- P. Natarajan, and R. Nevatia. (2007). Coupled hidden semi Markov models for activity recognition. In IEEE workshop on motion and video computing. WMVC '07.
- Q. Chen, Z. Song, Y. Hua, Z. Huang, and S. Yan (2012). Hierarchical matching with side information for image classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE International Conference on Computer Vision (ICCV).
- S. Fine, , Y. Singer, and N. Tishby. (1998). The hierarchical hidden Markov model: Analysis and applications. Machine Learning. 32(1): 41–62.
- S. Hongeng, R.Nevatia, and F. Bremond. (2004). Video-based event recognition: Activity representation and probabilistic recognition methods. Computer Vision and Image Understanding: CVIU, 96(2): 129–162.
- S. Park and J.K. Aggarwal. (2004). A hierarchical Bayesian network for event recognition of human actions and interactions. Multimedia Systems. 10(2): 164–179.
- S.Park, and J.K. Aggarwal. (2004). Semantic-level understanding of human actions and interactions using event hierarchy. In Proceedings of the 2004 conference on computer vision and pattern recognition workshop.
- Sun Ting and Geng Guohua. (2016). Image Retrieval Method for Deep Neural Network. international Journal of Signal Processing. Image Processing and Pattern Recognition. 9(7):33-42.
- Tianshui Chen, Muxin Xu, Xiaolu Hui, Hefeng Wu and Liang Lin. (2019). Learning Semantic-Specific Graph Representation for Multi-Label Image Recognition Computer Vision and Pattern Recognition (ICCV 2019).
- V. Joumier, E. Mulin, J. H Lee, J. Piano, A. Derreumaux, R. David, P. Mallea, A. Dechamps, P. H Robert, R. Romdhane, M. Thonnat and F. Bremond. (2012). Measurement instrument for assessing functional abilities of elderly people with and with- out dementia using a video monitoring system. In the 6th International Conference on Bioinformatics and Biomedical Engineering (ICBBE). 1-4.

- V.-T. Vu, Brémond, F., and Thonnat, M., (2003). Automatic video interpretation: A novel algorithm for temporal scenario recognition. Proceedings of the eighteenth international joint conference on artificial intelligence.1295–1302.
- Xiaodan Liang, Hongfei Zhou, Eric Xing. (2018). Dynamic-structured Semantic Propagation Network Computer Vision and Pattern Recognition.
- Y. A. Ivanov and A. F. Bobick. (2008). Recognition of visual activities and interactions by stochastic parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(8): 52–872.
- Y. Tsuruoka, J. Tsujii, S. Ananiadou. (2009). Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty, In Proceedings of the Asian Federation of Natural Language Processing.
- Z. Song, Q. Chen, Z. Huang, Y. Hua, and S. Yan (2011). Contextualizing object detection and classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Zinger, S., Millet, C., Mathieu, B., Grefenstette, G., Hede, P., and Mokklic, P-A. (2005). Extracting an ontology of portrayable objects from Wordnet. In Proceedings of the MUSCLE/ ImageCLEF Workshop on Image and Video Retrieval Evaluation.17-23.



ประวัติผู้วิจัย

รศ.นัทพ์ชาณัน ชินปัญช์ธนะ

มหาวิทยาลัยธุรกิจบัณฑิต 110/1-4 ถ . ประชาชน หลักสี่ กรุงเทพฯ 10210

โทร. 0-2954-7300 ต่อ 207 มือถือ 0-84-019-4484

Email: nutchanun.cha@dpu.ac.th

ปัจจุบัน

- อาจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิต
- คณะกรรมการกลุ่มผลิตชุดวิชาสาขา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
- คณะกรรมการ/ผู้เขียน
- การโปรแกรมบนเว็บ (Java Web Application) (ชุดวิชา 99420)
- การออกแบบส่วนปฏิสัมพันธ์บนเว็บและโมบาย (Web and Mobile Interaction Design) (ชุดวิชา 99319)
- การวิเคราะห์และออกแบบระบบเชิงวัตถุ (ชุดวิชา 99316) ม.สุโขทัยธรรมาธิราช (ผู้เขียน)

การศึกษา

พ.บ. สถิติประยุกต์ (คอมพิวเตอร์)สถาบันบัณฑิตพัฒนบริหารศาสตร์

วทบ. วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)มหาวิทยาลัยหอการค้าไทย

ได้รับทุนการศึกษาประเภทเรียนดี มหาวิทยาลัยหอการค้าไทย

ได้รับปริญญาตรีเกียรตินิยมอันดับ 1 จาก มหาวิทยาลัยหอการค้าไทย

สาขาวิจัย Image Retrieval, Image Processing, Classification and Recognition, Semantic Images, Human Activity

Certificate of Achievements

- นักวิจัยดีเด่นสาขาวิทยาศาสตร์ จาก มหาวิทยาลัยธุรกิจบัณฑิต 2558
- รับรองคุณภาพการสอนแบบก้าวหน้า (Progressive) จาก มหาวิทยาลัยธุรกิจบัณฑิต 2558
- Microsoft Office Specialist (MOS) (Excel2010) 2017
- Information Technology Professional Examination (ITPE) 2015
- The best paper award, The 8th National Conference on Applied Computer Technology and Information Systems, 2558.
- Internet and Computer Certification (IC3) 2014
- Microsoft Office Specialist (MOS) (Powerpoint2010, Word2010) 2014

- iOS Development (June 2013)

Book Chapter

- นัศพ์ชาณัณ ชินปัญชัณนะ หน่วยที่ 2 เลขฐาน, คอมพิวเบ็องติน (96101), มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2559.
- นัศพ์ชาณัณ ชินปัญชัณนะ, หน่วยที่ 4 การวิเคราะห์ความต้องการของระบบสารสนเทศ, การพัฒนาระบบสารสนเทศ (96407), มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2558.
- นัศพ์ชาณัณ ชินปัญชัณนะ หน่วยที่ 7 การจัดการแฟ้มข้อมูล, สถาปัตยกรรมคอมพิวเตอร์ (99315) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2556.
- นัศพ์ชาณัณ ชินปัญชัณนะ หน่วยที่ 10 โครงสร้างแบบตินไม้ และ หน่วยที่ 11 โครงสร้างแบบกราฟ, โครงสร้างข้อมูลและขั้นตอนวิธี (Data Structures and Algorithm) (99314) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2556.
- นัศพ์ชาณัณ ชินปัญชัณนะ, หน่วยที่ 1 ความรู้เบ็องตินเกี่ยวกับเครือข่ายคอมพิวเตอร์ และ หน่วยที่ 3 ชุมชนเครือข่าย, หลักการบริหารและจัดการเครือข่าย (Network Management) (99412) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2556.
- นัศพ์ชาณัณ ชินปัญชัณนะ, หน่วยที่ 12 โลจิสตีกส์และโซ่อุปทานในระบบพาณิชย์อิเล็กทรอนิกส์ และ หน่วยที่ 13 การชำระเงินในระบบพาณิชย์อิเล็กทรอนิกส์, ระบบสานักงานอัตโนมัติและพาณิชย์อิเล็กทรอนิกส์ (Office Automation System and Electronic Commerce) (99311) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิทยาศาสตร์และเทคโนโลยี 2555.
- นัศพ์ชาณัณ ชินปัญชัณนะ หน่วยที่ 3 โพรโทคอลล และ หน่วยที่ 5 เอ็กซ์เอ็มแอล, เทคโนโลยีการบริการผ่านเว็บและการประยุกต์ (Web Services Technology and Applications) (99301) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2555.
- นัศพ์ชาณัณ ชินปัญชัณนะ, หน่วยที่ 3 ตั๋วกลางในการสื่อสารข้อมูลและ หน่วยที่ 10 โพรโทคอลล, การสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์ (Data Communications and Networking) มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขา วิทยาศาสตร์และเทคโนโลยี 2553.

รางวัล/Awards

- รับรางวัลรองชนะเลิศ รางวัลรองชนะเลิศอันดับ 2 แอปพลิเคชัน Lawang-chang System เรื่องระบบการตรวจจับและแจ้งเตือนข้างออกมานอกพื้นที่ จาก การแข่งขัน G-Con 2018, พฤษภาคม 2561 (อาจารย์ที่ปรึกษา)

- รับรางวัลรองชนะเลิศ ด้านความมั่นคง เรื่องระบบการตรวจสอบสิ่งกีดขวางในแนวท่อน้ำมัน ใน นำเสนอและการประกวด อากาศยานไร้คนขับ (Unmanned Aerial Vehicle : UAV)” ประจำปี ๒๕๖๐ จากสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) เมื่อวันที่ 12-13 สิงหาคม 2560.
- รับรางวัล The best paper เรื่อง “การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟ แบบจับคู่.” งานประชุมทางวิชาการระดับชาติพะเยาวิจัย ครั้งที่ 6 วิทยาศาสตร์ เทคโนโลยี และ นวัตกรรม เพื่อความเข้มแข็งของชุมชน , มหาวิทยาลัยพะเยา, จังหวัดพะเยา, มกราคม 2560. (The best paper award)
- รับรางวัล The best paper เรื่อง , "การคัดเลือกข้อมูลเพื่อใช้จำแนกโครงสร้างท่าทางมนุษย์ด้วยซอฟต์แวร์เวกเตอร์แมชชีน" จาก The 8th National Conference on Applied Computer Technology and Information Systems 2015. (The best paper award)
- ผ่านโครงการสอบมาตรฐานวิชาชีพไอที (ITPE) สถาบันวิทยากร สวทช. ผ่านระดับ 1: IP ครั้งที่ 2/2558 <https://www.nstdaacademy.com/webnsa/index.php/home/all-news/1078-itpe-sss-result0258>
- รับรองคุณภาพการสอน ปีการศึกษา 2557 จาก มหาวิทยาลัยธุรกิจบัณฑิตย์ (เทอม 1/2558)
- รับรางวัลนักวิจัยดีเด่น สาขาเทคโนโลยีสารสนเทศ 2558 จาก มหาวิทยาลัยธุรกิจบัณฑิตย์
- Certificate of Achievement for iOS Development 17-21 June 2013
- Microsoft Office Specialist (MOS) Certification (Excel 2013, Year 2017)
- Microsoft Office Specialist (MOS) Certification (PowerPoint and Word 2010, Year 2015)

การแข่งขันประกวด

- การประกวด อากาศยานไร้คนขับ (Unmanned Aerial Vehicle : UAV)” รางวัลรองชนะเลิศอันดับที่ 2 เรื่องการตรวจจับวัตถุต้องสงสัยบนเส้นทางด้วยการประมวลผลภาพ ประจำปี ๒๕๖๐ จาก สำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) เมื่อวันที่ 12-13 สิงหาคม 2560.
- การแข่งขัน G-Con 2018 รางวัลรองชนะเลิศ เรื่องการแจ้งเตือนข้างออกนอกพื้นที่, พฤษภาคม 2561

งานบริการทางวิชาการ

- Reviewer Committee: Imaging Science Journal 2019 <https://www.editorialmanager.com/ims/default.aspx>
- Reviewer Committee: a program committee of The Seventh ICT International Student Project Conference (ICT-ISPC 2018).
- Reviewer Committee: a program committee of ICLIST and NCLIST 2018 program committee the conference on March 21 – 24, 2018 at Novotel Hua Hin Cha Am Beach Resort Thailand.
- Reviewer Committee: The 15th International Joint Conference on Computer Science and Software Engineering (JCSSE2018), Thailand, 2018.

- Reviewer Committee: European Journal of Remote Sensing on Apr-2018
- คณะกรรมการพิจารณาบทความ : การประชุมวิชาการด้านเทคโนโลยีสารสนเทศ ครั้งที่ 9 (NCIT 2017) วันที่ 1-2 พ.ย. 2560 อำเภอพุทธมณฑล จังหวัดนครปฐม
- Reviewer Committee:, 5th International Conference on Advanced Computer Science Applications and Technologies (ACSAT2017), Kuala Lumpur-Malaysia, 2017.
- Reviewer Committee: International Conference on Mechatronics Systems and Control Engineering (ICMSCE 2017), Amsterdam, 2017.
- คณะกรรมการพิจารณาบทความ : การประชุมวิชาการด้านเทคโนโลยีสารสนเทศ ครั้งที่ 6 (NCIT 2014) วันที่ 27-28 ก.พ. 2557 ณ ทองสมบูรณ์ คลับ จ.นครราชสีมา
- คณะกรรมการกลุ่มผลิตชุดวิชาการสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์ 1/2552 สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
- คณะกรรมการกลุ่มผลิตชุดวิชาการบริการผ่านเว็บ 1/2553 สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อบรมวิทยากร

- วิทยากรบรรยาย เรื่อง AI technology for Image Processing Session 3.1: Introduction to Image Processing (DataTech) ต.ค.2018
- วิทยากรบรรยาย เรื่อง การพัฒนากระบวนการคิดเชิงบูรณาการ STEM EDUCATION วิทยากรบรรยาย ณ อาคาร 30 คณะครุศาสตร์ มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา วันที่ 9-10 มิ.ย. 2561.
- วิทยากรบรรยาย ฝึกปฏิบัติ Google Application: Google Site กับคุณครู ณ โรงเรียนหอวังนนทบุรี พ.ค. 2561.
- วิทยากรบรรยาย ฝึกปฏิบัติ เรื่อง การจัดการศึกษาด้วยการบูรณาการความรู้ด้วยหุ่นยนต์ (STEM Robotic Education) กับครูกศน.กรุงเทพมหานคร 20-21 พ.ย. 2560
- วิทยากรบรรยาย ฝึกปฏิบัติโปรแกรมภาษาจาวาขั้นต้น วันที่ 18-20 ตุลาคม 2556 อาคารเฉลิมพระเกียรติ มหาวิทยาลัยสุโขทัยธรรมาธิราช
- วิทยากรบรรยาย โครงการฝึกอบรมหลักสูตร การพัฒนาทักษะด้านคอมพิวเตอร์ สำนักงานกองทุนให้กู้ยืมเพื่อการศึกษา รุ่นที่ 1 และ รุ่นที่ 2 (Microsoft Access 2010) วันที่ 18 สิงหาคม 2556 และ วัน 1 กันยายน 2556 มหาวิทยาลัยธุรกิจบัณฑิตย์
- วิทยากรบรรยายในการผลิตสื่อ e-learning วันที่ 9 กรกฎาคม 2556 ประกอบชุดวิชา 99412 หลักการและการบริหารเครือข่าย เรื่อง ความรู้เบื้องต้นเกี่ยวกับระบบเครือข่ายคอมพิวเตอร์ และเรื่องสังคมระบบเครือข่าย สำนักเทคโนโลยีการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช
- วิทยากรบรรยาย Microsoft Office PowerPoint and Microsoft Visio 2010 วันที่ 3 มีนาคม 2556 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย
- วิทยากรบรรยาย Tip เด็ดๆจากการใช้งาน Ms.Word 2010)บุคคลากรภายใน(วันที่ 12 มีนาคม 2558 มหาวิทยาลัยธุรกิจบัณฑิตย์
- วิทยากรบรรยาย Tip เด็ด Ms.Word 2010)บุคคลากรภายใน(วันที่ 7 และ 16 กรกฎาคม 2557 มหาวิทยาลัยธุรกิจบัณฑิตย์
- วิทยากรบรรยาย Microsoft Powerpoint 2010)บุคคลากรภายนอก คณะครู จังหวัดสระบุรี(วันที่ 29 ตุลาคม 2556 มหาวิทยาลัยธุรกิจบัณฑิตย์

- วิทยากรบรรยาย ฝึกปฏิบัติโปรแกรมภาษาจาวาขั้นต้น วันที่ 18-20 ตุลาคม 2556 อาคารเฉลิมพระเกียรติ มหาวิทยาลัยสุโขทัยธรรมาธิราช
- วิทยากรบรรยาย โครงการฝึกอบรมหลักสูตร การพัฒนาทักษะด้านคอมพิวเตอร์ สำนักงานกองทุนให้กู้ยืมเพื่อการศึกษา รุ่นที่ 1 และ รุ่นที่ 2)Microsoft Access 2010) วันที่ 18 สิงหาคม 2556 และ วัน 1 กันยายน 2556 มหาวิทยาลัยธุรกิจบัณฑิต
- วิทยากรบรรยายในการผลิตสื่อ e-learning วันที่ 9 กรกฎาคม 2556 ประกอบชุดวิชา 99412 หลักการและการบริหารเครือข่าย เรื่อง ความรู้เบื้องต้นเกี่ยวกับระบบเครือข่ายคอมพิวเตอร์ และเรื่องสังคระบบเครือข่าย สำนักเทคโนโลยีการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช
- วิทยากรบรรยาย Microsoft Office PowerPoint and Microsoft Visio 2010 วันที่ 3 มีนาคม 2556 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย
- วิทยากรบรรยาย ฝึกปฏิบัติโปรแกรมภาษาจาวาขั้นต้น ตุลาคม 2555 อาคารเฉลิมพระเกียรติ มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์พิเศษ

มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี วิชาโครงสร้างข้อมูล (Data Structure)) 1/2541) (Tool: Borland Turbo C Programming language)

มหาวิทยาลัยศรีปทุม วิชาโครงสร้างข้อมูล (Data Structure) (Tool: C++ program)

มหาวิทยาลัยเกริก วิชาคลังข้อมูล (Data warehouse)

วิทยาลัยภาคกลาง จ.นครสวรรค์ วิชาความปลอดภัยข้อมูล(Computer Security)

หน่วยที่ 1. ความเบื้องต้นเกี่ยวกับเครือข่ายคอมพิวเตอร์ ผู้แต่ง ผศ.นัศพชาณัณ ชินปัญชธนะ

หน่วยที่ 3. ชุมชนเครือข่าย ผู้แต่ง ผศ.นัศพชาณัณ ชินปัญชธนะ

Research

- นัศพชาณัณ ชินปัญชธนะ, เรื่องการจัดกลุ่มความหมายภาพส่วนบุคคลด้วยข้อความกิจกรรมเหตุการณ์ (Semantic Clustering for Personal Image with Context of Activity Events) มหาวิทยาลัยธุรกิจบัณฑิต, 2557.
- นัศพชาณัณ ชินปัญชธนะ, การแปลความหมายภาพด้วยแนวคิดพื้นฐานความสัมพันธ์ของกราฟแบบลำดับชั้น (Semantic Annotation Model of Hierarchical Relationships based on Conceptual Graph Representation), มหาวิทยาลัยธุรกิจบัณฑิต, 2556.
- นัศพชาณัณ ชินปัญชธนะ, การสุจาทาทางมนุษย์จากภาพส่วนบุคคลผ่านการวิเคราะห์การใช้พลังงานร่างกาย (Human Action Recognition from Personal Photos via Analysis of Body Energy Expenditure), มหาวิทยาลัยธุรกิจบัณฑิต, 2555.
- นัศพชาณัณ ชินปัญชธนะ, การแปลความหมายภาพด้วยวิธีวัดความคล้ายกันของกราฟแบบจับคู่, มหาวิทยาลัยธุรกิจบัณฑิต, 2554.
- นัศพชาณัณ ชินปัญชธนะ, ระบบตรวจนับวัตถุอัตโนมัติด้วยเทมเพลตแมชชีนซึ่งแบบนอร์มอลไลซ์คอร์รีเลชัน (Automatic Counting Objects System by using Template Matching with Normalized Correlation) 2553.

- นัศพ์ชาณัณ ชินปัญช์ธนะ, การจําแนกความหมายของภาพจากวัตถุโดยใช้หลักการโครงสร้างสเกตริตรอน, มหาวิทยาลัยธุรกิจบัณฑิต, 2552.

Publications

- นัศพ์ชาณัณ, ชินปัญช์ธนะ การศึกษาเทคนิคงานวิจัยทางการประมวลผลภาพดิจิทัล, การประชุมวิชาการระดับชาติ ครั้งที่ 14 Digital Technology for Sustainable Wellbeing and Smart Society มหาวิทยาลัยธุรกิจบัณฑิต, 2562.
- นัศพ์ชาณัณ ชินปัญช์ธนะ, “การจําแนกภาพความสัมพันธ์ของ แอ็คชั่นด้วยวิธีการโครงข่ายประสาทเทียมแบบสังวัตนาการซัพพอร์ตเวกเตอร์แมชชีน,” วิศวกรรมสารเกษมบัณฑิต ปีที่ 9 ฉบับที่ 1 มกราคม-เมษายน 2562, หน้า 186-203.
- นัศพ์ชาณัณ ชินปัญช์ธนะ,การจําแนกความหมายภาพด้วยความสัมพันธ์แอ็คชั่นภายในเพื่อช่วยเหลือผู้สูงอายุ,การประชุมวิชาการระดับชาติ “วลัยลักษณ์วิจัย” ครั้งที่ 11,นครศรีธรรมราช,2562.
- ณัฐวัฒน์ คำเคน,กิตติพงษ์ โพธิ์เป้า, นัศพ์ชาณัณ ชินปัญช์ธนะ, อุดมลักษณ์ อําพันธู์และสํารายู ใฝ่นवल, แอปพลิเคชันร้านขายไก่อํางสำหรับธุรกิจออนไลน์,การประชุมวิชาการระดับชาติ NCTIM2019, มหาวิทยาลัยราชภัฏมหาสารคาม,มหาสารคาม,2562.
- ดารณั ตั้งสุทธิมงคล,งามพล เทียนทอง,นัศพ์ชาณัณ ชินปัญช์ธนะ, อุดมลักษณ์ อําพันธู์และสํารายู ใฝ่นवल,แอปพลิเคชันออนไลน์สําหรับโรงเรียนสอนขับรถตามตารางนัดหมาย,การประชุมวิชาการระดับชาติ NCTIM2019, มหาวิทยาลัยราชภัฏมหาสารคาม,มหาสารคาม,2562.
- นัศพ์ชาณัณ ชินปัญช์ธนะ, อุดมลักษณ์ อําพันธู์และสํารายู ใฝ่นवल, การศึกษาวิธีการหาความหมายภาพด้วยเทคนิคการค้นคืนคุณภาพดิจิทัล, งานประชุมวิชาการระดับชาติ NCTIM2019, มหาวิทยาลัยราชภัฏมหาสารคาม,มหาสารคาม,2562.
- Nuchanun Chinpanthana and Tejtasin Phiasai, Spatial Semantic Images with Relationship Contents by using Convolutional Neural Network and Support Vector Machine, 2018 VII International Conference on Network, Communication and Computing, in Howard Civil Service International House on December 14-15, 2018 (published by ACM (ISBN: 978-1-4503-6553-6))
- นัศพ์ชาณัณ ชินปัญช์ธนะ, การหาความหมายภาพด้วยการวัดความคล้ายภาพด้วยเปรียบเทียบคู่ที่เหมาะสม,ประชุมวิชาการวิจัยและนวัตกรรมสร้างสรรค์ ครั้งที่ 5 หัวข้อ “สู่วิจัยรับใช้สังคม ด้วยนวัตกรรมสร้างสรรค์” CRCI-2018 ณ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ตาก ระหว่าง วันที่ 6 – 8 ธันวาคม พ.ศ. 2561
- นัศพ์ชาณัณ ชินปัญช์ธนะ, สํารายู ใฝ่นवल, เมธัส ชุมเจริญ, สุทธิศักดิ์ เจริญผล และ ตินณภัทร์ ศรีสุขสมบูรณ์, แอปพลิเคชันแจ้งเตือนเพื่อตรวจจับกิจกรรมต้องสงสัยในเขตป่าไม้, NCLIST ณ โรงแรม โนโวเทล หัวหินชะอำ ปีช รีส์อร์ท จ.เพชรบุรี ระหว่างวันที่ 21-24 มีนาคม 2561
- อุดมลักษณ์ อําพันธู์, ณรงค์ฤทธิ ดวงประทุมข และ นัศพ์ชาณัณ ชินปัญช์ธนะ, แอปพลิเคชันแจ้งเตือนข้างออกนอกพื้นที่ป่าไม้แบบเรียลไทม์, NCLIST ณ โรงแรม โนโวเทล หัวหิน ชะอำ ปีช รีส์อร์ท จ.เพชรบุรี ระหว่างวันที่ 21-24 มีนาคม 2561

- นศพ์ชาณัณ ชินปัญชธนะ การคั่นคีนข้อมูลภาพด้วยการทำนายความสัมพันธ์ของแอ็คชั่น การประชุมทางวิชาการระดับชาติพะเยาวิจัย ครั้งที่ 7, มหาวิทยาลัยพะเยา, วันที่ 25-26 มกราคม 2561
- Kawisara Boonjun, Narongdech Keeratipranon and Nutchanut Chinpanthana, “Real time Automatic Object Detection by using Template Matching for Protecting Pipelines,” International Workshop on Advanced Image Technology 2018 (IWAIT 2018) January 7-10, 2018 in Chiang Mai, Thailand.
- N. Chinpanthana and T. Phiasai, “Deep Textual Searching for Visual Semantics of Personal Photo Collections with a Hybrid Similarity Measure,” 2017 International Symposium on Computer Science and Intelligent Controls, Oct 20-22, Hungary, 2017.
- นศพ์ชาณัณ ชินปัญชธนะ, การคั่นคีนฐานข้อมูลภาพดิจิทัลด้วยความสัมพันธ์แบบกราฟลำดับชั้น,”การประชุมวิชาการระดับชาติและนานาชาติ ราชภัฏสุราษฎร์ธานีวิจัย ครั้งที่ 13 ในวันที่ 22 - 24 พฤศจิกายน 2560.
- นศพ์ชาณัณ ชินปัญชธนะ, “การคั่นคีนภาพดิจิทัล ด้วยการวัด ความคล้ายกันบนกราฟแบบมีทิศทางหลายลำดับ,” วิศวกรรมสารเกษมบัณฑิต ปีที่ 7 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2560, 60-75. (วารสารอยู่ในฐานข้อมูล TCI กลุ่มที่1 ISSN 2286-668X)
- นศพ์ชาณัณ ชินปัญชธนะและ เตชศุทธิณบ์ เพี้ยชัย, “ระบบนับวัตถุอัตโนมัติด้วยเทมเพลตแมชชีนแบบฟาสเตอร์ มัลโลกอร์สคอร์เรชัน,” วารสารวิทยาศาสตร์และเทคโนโลยี มทร.ธัญบุรี, ปีที่ 7, ฉบับที่ 2, กรกฎาคม-ธันวาคม, 2560, 168-182. (วารสารอยู่ในฐานข้อมูล TCI กลุ่มที่1 ISSN 2229-1547)
- นศพ์ชาณัณ ชินปัญชธนะ, “ศึกษาการสกัดคุณลักษณะข้อมูลสำหรับระบบคั่นคีนข้อมูลภาพ,” วารสาร มหาวิทยาลัยคริสเตียน, 2560.
- นศพ์ชาณัณ ชินปัญชธนะ, “การวัดความคล้ายกันของภาพด้วยการเชื่อมโยงกราฟแบบลำดับชั้นบนพื้นฐานของอนุกรมวิธาน WordNet,” The Journal of Information Science and Technology (JIST), Vol 6, No. 1,(NCIT2017 selected paper), 2017.
- N. Chinpanthana, “Personal Image Retrieval with Hierarchical Similarity Measure Based on WordNet Framework” 2017 International Conference on Materials Engineering and Functional Materials (ICMFM 2017) Hanoi, Vietnam, 2017.
- อนุวัช พรภิไหว,นศพ์ชาณัณ ชินปัญชธนะ และนันทิกา ปริญญาพล, “การพัฒนาแอปพลิเคชันบนมือถือเพื่อค้นหาพื้นที่เพาะปลูกข้าวและพื้นที่ป่าไม้ในประเทศไทยด้วยภาพถ่ายดาวเทียมและระบบสารสนเทศภูมิศาสตร์,” งานประชุมวิชาการระดับชาติ วิทยาศาสตร์และเทคโนโลยีระหว่างสถาบัน ครั้งที่5, โรงแรมมิราเคิลแกรนด์ คอนเวนชั่น, กรุงเทพฯ 2560.
- นศพ์ชาณัณ ชินปัญชธนะ และ สารานู ใฝ่นวล, การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่, งานประชุมทางวิชาการระดับชาติพะเยาวิจัย ครั้งที่ 6 “วิทยาศาสตร์ เทคโนโลยี และนวัตกรรมเพื่อความเข้มแข็งของชุมชน”, มหาวิทยาลัยพะเยา, จังหวัดพะเยา, มกราคม 2560. (บทความดีเด่น กลุ่มวิทยาศาสตร์)
- นศพ์ชาณัณ ชินปัญชธนะ และเตชศุทธิณบ์ เพี้ยชัย, ระบบตรวจจับวัตถุอัตโนมัติด้วยเทมเพลตแมชชีนแบบนอร์มัลไลซ์คอร์เรชัน” การประชุมวิชาการระดับชาติและนานาชาติ ราชภัฏสุราษฎร์ธานีวิจัย ครั้งที่ 12, มหาวิทยาลัยราชภัฏสุราษฎร์ธานี, ตุลาคม,2559.

- นัศพ์ชาณัณ ชินปัญชัณนะ, การวัดความคล้ายกันของภาพด้วยการเชื่อมโยงกราฟแบบลำดับชั้นบนพื้นฐานของอนุกรมวิธาน WordNet,การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ (NCIT),2559.
- นัศพ์ชาณัณ ชินปัญชัณนะ และ สำราญ ไผ่นวล, “การแปลความหมายภาพด้วยวิธีการวัดความคล้ายกันของกราฟแบบจับคู่,” งานประชุมทางวิชาการระดับชาติพะเยาวิจัย ครั้งที่ 6 วิทยาศาสตร์ เทคโนโลยี และนวัตกรรม เพื่อความเข้มแข็งของชุมชน , มหาวิทยาลัยพะเยา, จังหวัดพะเยา, มกราคม 2560. (The best paper award)
- N. Chinpanthana, “Semantic Clustering Based on Context of Activity Events in Personal Photo Collections,” International Workshop on Smart Info-Media Systems in Asia, Chiba Institute of Technology Narashina, Japan, 2015.
- นัศพ์ชาณัณ ชินปัญชัณนะ, สำราญ ไผ่นวล และ ริญญรัตน์ โชติสุริยสินสุข, “การศึกษางานวิจัยการประมวลผลภาพดิจิทัลและการประยุกต์ใช้งานในแอปพลิเคชัน,” การประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏเพชรบูรณ์ ครั้งที่ 3, 2559.
- วิจารณ์ ประดับ, นัศพ์ชาณัณ ชินปัญชัณนะ และ ริญญรัตน์ โชติสุริยสินสุข, “ทิศทางการวิจัยของการแปลความหมายภาพด้วยการแบ่งส่วนข้อมูลภาพ,” การประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏนครปฐม ครั้งที่ 8, จังหวัดนครปฐม, 31 มีนาคม - 1 เมษายน 2559.
- นัศพ์ชาณัณ ชินปัญชัณนะ, “การจัดกลุ่มความหมายภาพเหตุการณ์กิจกรรมด้วยพีชชีซีมีน” งานประชุมทางวิชาการระดับชาติพะเยาวิจัย ครั้งที่ 5 “วิทยาศาสตร์ เทคโนโลยี และนวัตกรรม เพื่อความเข้มแข็งของชุมชน”, มหาวิทยาลัยพะเยา, จังหวัดพะเยา, มกราคม 2559.
- นัศพ์ชาณัณ ชินปัญชัณนะ, “ทิศทางการวิจัยของงานวิจัยทางด้านการประมวลผลภาพดิจิทัลในการแปลความหมายภาพ, ในงานประชุมวิชาการระดับชาติและนานาชาติมหาวิทยาลัยราชภัฏอุตรดิตถ์ ครั้งที่ 11, มหาวิทยาลัยราชภัฏอุตรดิตถ์, จังหวัดอุตรดิตถ์, ธันวาคม 2558.
- นัศพ์ชาณัณ ชินปัญชัณนะ, “การจำแนกความหมายภาพด้วยการสกัดพีเจอร์จากโครงสร้างสเกตริตรอนบนพื้นฐานแนวคิดกราฟลำดับชั้น”, วารสารแม่โจ้เทคโนโลยีสารสนเทศและนวัตกรรม,1, 2558.
- นัศพ์ชาณัณ ชินปัญชัณนะ, “การคัดเลือกข้อมูลเพื่อใช้จำแนกโครงสร้างท่าทางมนุษย์ด้วยซอฟต์แวร์คอมพิวเตอร์แมชชีน”, “The 8th National Conference on Applied Computer Technology and Information Systems”, วันที่ ๓๐ - ๓๑ มกราคม ๒๕๕๘, มหาวิทยาลัยนครพนม,จังหวัดนครพนม. (The best paper award)
- นัศพ์ชาณัณ ชินปัญชัณนะ, “การจำแนกความหมายภาพด้วยเรเดียลเบสิสฟังก์ชันบนพื้นฐานของแนวคิดกราฟลำดับชั้น” ,นานาชาติการวิจัยและนวัตกรรมเพื่อการพัฒนาที่ยั่งยืนสู่เอเชีย” วันศุกร์ที่ 10 ตุลาคม 2557 , วิทยาลัยบัณฑิตเอเชีย จังหวัดขอนแก่น
- N. Chinpanthana and T. Phiasai., “Kernel-based on Data Fusion for Image Classification with Body Energy Action Model”, International Journal of Signal Processing Systems, vol. 1, no. 2, December, 2013.
- N. Chinpanthana and T. Phiasai., “Kernel-based on Data Fusion for Image Classification with Body Energy Action Model”, The 2013 5th International Conference on Signal Processing Systems (ICSPS 2013), Sydney, Australia, 2013.

- 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”, International 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”, International Workshop on Smart Info-Media Systems, Thailand, 2007.
- 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,” International 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.