

การติดตามหลากหลายเป้าหมายโดยใช้การเชื่อมโยงข้อมูล
และการปรับเรียบเส้นทางการเดิน

MULTIPLE TARGET TRACKING USING DATA ASSOCIATION
AND TRAJECTORY SMOOTHING



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาคามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2561

KMITL-2018-EN-M-070-084

การติดตามหลากหลายเป้าหมายโดยใช้การเชื่อมโยงข้อมูล
และการปรับเรียบเส้นทางการเดิน

MULTIPLE TARGET TRACKING USING DATA ASSOCIATION
AND TRAJECTORY SMOOTHING



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2561

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MULTIPLE TARGET TRACKING USING DATA ASSOCIATION
AND TRAJECTORY SMOOTHING



CHAIPITAK PATTANAKITKHUN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN COMPUTER ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2018
KMITL-2018-EN-M-070-084

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2018

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การติดตามหลากหลายเป้าหมายโดยใช้การเชื่อมโยงข้อมูลและการปรับเรียบ
เส้นทางการเดิน

Thesis Title Multiple Target Tracking using Data Association and Trajectory Smoothing

นักศึกษา นายชัยพิทักษ์ พัฒนภิตติคุณ

รหัสประจำตัว 58601099

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.อรฉัตร จิตต์โสภักตร์

หมายเลขวิทยานิพนธ์ KMITL-2018-EN-M-070-084

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
รศ.ดร.สุรพงศ์	เอื้อวัฒนามงคล	อรฉัตร จิตต์โสภักตร์
รศ.ดร.เกียรติกุล	เจียรนัยธนะกิจ	คส
ผศ.ดร.สมศักดิ์	วัลย์รัชต์	ค
รศ.ดร.บุญธีร์	เครือตราชู	กมฉ
รศ.ดร.อรฉัตร	จิตต์โสภักตร์	อรฉัตร จิตต์โสภักตร์

วัน / เดือน / ปี ที่สอบ วันจันทร์ที่ 9 กรกฎาคม พ.ศ. 2561 เวลา 09.30-11.30 น.
สถานที่สอบ ณ ห้อง HM-301 อาคารเฉลิมพระเกียรติ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิศวกรรมศาสตร์ รับรองแล้ว

(รองศาสตราจารย์ ดร. คมตัน มาลีสี)

คณบดี คณะวิศวกรรมศาสตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่ในช่องทางอื่น
วันที่ 9 กรกฎาคม พ.ศ. 2561
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การติดตามหลากหลายเป้าหมายโดยใช้การเชื่อมโยงข้อมูลและการปรับเรียงเส้นทางการเดิน
นักศึกษา	นายชัยพิทักษ์ พัฒนภิตติคุณ
รหัสประจำตัว	58601099
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2561
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.อรฉัตร จิตต์โสภักตร์

บทคัดย่อ

การติดตามหลากหลายวัตถุประสงค์ นั้นเป็นงานวิจัยที่สำคัญยิ่งในวงการของคอมพิวเตอร์วิทัศน์ วิธีการที่เป็นที่นิยมชนิดหนึ่งคือการเชื่อมโยงข้อมูลโดยพิจารณาจากทุกเฟรมตลอดทั้งวิดีโอพร้อมกัน โดยจะสร้างแตรัคเก็ตจากช่วงระยะเวลาสั้นๆก่อน และทำการรวมแตรัคเก็ตเป็นเส้นทางการเดินที่สมบูรณ์ ดังนั้นงานวิจัยนี้จึงนำเสนอการปรับปรุงการเชื่อมโยงข้อมูลเพื่อให้ได้ผลลัพธ์เส้นทางการเดินที่มีประสิทธิภาพมากยิ่งขึ้น ซึ่งประกอบด้วย 2 ขั้นตอนที่สำคัญ ได้แก่ ขั้นตอนการสร้างเส้นตรงตัวแทนของแตรัคเก็ตและปรับปรุงประสิทธิภาพของแตรัคเก็ตด้วยโหนดจากเส้นตรงตัวแทนของแตรัคเก็ต เพื่อให้เกิดความแม่นยำของแตรัคเก็ตมากยิ่งขึ้น และอีกขั้นตอนคือ การรวมแตรัคเก็ตเป็นเส้นทางการเดินและปรับปรุงความถูกต้องด้วยเทคนิคการปรับเรียงข้อมูล โดยจะทำการประมาณเส้นทางการเดินและทำการคัดเลือกแตรัคเก็ตที่มีค่าอยู่ในเกณฑ์ นามารมเป็นค ทอบเส้นทางการเดินที่สมบูรณ์ จากผลการทดสอบประสิทธิภาพพบว่า ค่าเฉลี่ยค่าความสมบูรณ์ของการติดตามของงานวิจัยที่นำเสนอมีค่าเท่ากับ 81.06% ซึ่งเพิ่มขึ้นถึง 8.17%

Thesis	Multiple target tracking using data association and trajectory smoothing
Student	Mr.Chaipitak Pattanakittikhun
Student ID	58601099
Degree	Master of Engineering
Program	Computer Engineering
Year	2018
Thesis Advisor	Assoc. Prof. Dr.Orachat Chitsobhuk

ABSTRACT

Multi Object tracking is an important task in the field of computer vision. One of the popular tracking methods is data association, which analyzes every image frames of the whole video sequence, generates multiple low-level tracklets, and merges them into trajectories. An improvement of the data association algorithm is proposed in this research. This proposed algorithm consists of two significantly processes. In the first process, the proposed algorithm tries to eliminate the outliers of tracklets and performs node prediction from extracted inliers in order to create the predicted nodes and lines. Tracklet is then generated from the integration of the extracted inliers of original tracklets and the results of node prediction. In the second process, the proposed algorithm merges tracklets into trajectories and performs trajectory improvement using smoothing data technique in order to eliminate the outlier tracklets. Only inlier tracklets are selected for merging to final trajectories. The performance of the proposed system is measured in terms of Track Completeness (TC). It can be seen that the proposed trajectory offers high accuracy at 81.06% and can help to improve the system performance by 8.17%.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงตามวัตถุประสงค์ได้ด้วยความรู้ของรองศาสตราจารย์ ดร.อรฉัตร จิตต์โสภักดิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ซึ่งให้คำปรึกษา ข้อชี้แนะ ตักเตือนและให้แนวทางในการแก้ปัญหาต่างๆ ทั้งด้านงานวิจัย ด้านวิชาการ ด้านการดำเนินชีวิต และให้ความช่วยเหลือในหลายสิ่งหลายอย่างแก่ลูกศิษย์คนนี้ หากไม่มีคำชี้แนะจากอาจารย์ วิทยานิพนธ์ฉบับนี้คงสำเร็จลุล่วงไปไม่ได้ ข้าพเจ้าขอกราบขอบพระคุณท่านอาจารย์เป็นอย่างสูงมา ณ ที่นี้

ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาและครอบครัวที่ให้อำนาจใจข้าพเจ้าจนสามารถทำวิทยานิพนธ์เล่มนี้จนสำเร็จลุล่วง

ขอขอบคุณ พี่ๆ เพื่อนๆ และน้องในภาควิชาวิศวกรรมคอมพิวเตอร์ทุกคน ที่คอยสอบถาม ให้ความช่วยเหลือ ให้คำแนะนำและกำลังใจที่ดีมาตลอด

และที่สำคัญ ขอขอบคุณความเพียรพยายามของตัวข้าพเจ้า

คุณความดีและประโยชน์ทั้งหลายที่ได้จากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบแต่บิดามารดา และครอบครัวซึ่งเป็นที่ยึดของข้าพเจ้า ตลอดจนครูบาอาจารย์ที่เคารพรักทุกท่านที่ประสิทธิ์ประสาทวิชาให้กับตัวข้าพเจ้า

ชัยพิทักษ์ พัฒนกิตติคุณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อ.....	I
บทคัดย่อ(ภาษาอังกฤษ).....	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา	1
1.3 สมมติฐานของการศึกษา.....	1
1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย	2
1.5 ขอบเขตงานวิจัย.....	2
1.6 ขั้นตอนการวิจัย.....	3
1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย.....	3
1.8 โครงสร้างของวิทยานิพนธ์	3
บทที่ 2 งานวิจัยที่เกี่ยวข้อง.....	5
2.1 การเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วน	5
2.2 การเชื่อมโยงข้อมูลโดยพิจารณาจากภาพรวม	7
บทที่ 3 ทฤษฎีพื้นฐานที่เกี่ยวข้อง.....	11
3.1 ค่าความถี่ของทิศเกรเดียนท์.....	11
3.1.1 ตัวแปรเริ่มต้นในการสกัดคุณลักษณะค่าความถี่ของทิศเกรเดียนท์	11
3.1.2 การคำนวณค่าเวกเตอร์ความลาดชัน ค่าขนาดการเปลี่ยนแปลงของความเข้มแสง และทิศทางของเกรเดียนท์เวกเตอร์.....	12
3.1.3 การคำนวณค่าความถี่ของทิศเกรเดียนท์ภายในเซลล์	13
3.1.4 นอร์มัลไลเซชันบล็อก.....	13
3.1.5 ค่าเวกเตอร์คุณลักษณะ	14
3.2 การขจัดค่าที่ไม่มากที่สุดสำหรับการตรวจจับวัตถุ	15
3.3 การติดตามหลากหลายวัตถุโดยใช้เทคนิคการพิจารณาตามเกณฑ์ของคลิกรูปที่สั้นที่สุด .	16
3.3.1 การหาแทร็คเล็กต์โดยวิธีการพิจารณาตามเกณฑ์ของคลิกรูปที่สั้นที่สุด	17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

3.3.2 การรวมแตรีกเล็ดเป็นเส้นทางการเดินโดยใช้การพิจารณาตามเกณฑ์ ของคลิกกราฟที่สั้นที่สุด	23
3.4 การถดถอยเชิงเส้น	27
3.5 การปรับเรียบข้อมูลแบบถดถอยโดยพิจารณาเฉพาะส่วน	28
3.5.1 สมการถดถอยโดยพิจารณาเฉพาะส่วน	28
3.5.2 สมการถดถอยโดยพิจารณาเฉพาะส่วนแบบแกร่ง	29
3.6 การวัดประสิทธิภาพ	31
3.6.1 การวัดประสิทธิภาพสำหรับการตรวจจับวัตถุ	31
3.6.1.1 การวัดประสิทธิภาพความถูกต้องและความผิดพลาดในการตรวจจับ ผลบวกจริง, ผลบวกลวง, ผลลบจริง, ผลลบลวง	31
3.6.1.2 ค่าความแม่นยำ	33
3.6.1.3 ค่าความครบถ้วน	33
3.6.1.4 ค่าความถูกต้องในการตรวจจับหลายวัตถุ	34
3.6.1.5 ค่าความแม่นยำในการตรวจจับหลายวัตถุ	34
3.6.2 การวัดประสิทธิภาพสำหรับการติดตามวัตถุ	35
3.6.2.1 ค่าความสมบูรณ์ของการติดตาม	35
3.6.2.2 ค่าระยะที่ผิดพลาดของการติดตาม	35
3.6.2.3 การวัดระยะแบบยูคลิเดียน	36
3.6.2.4 ค่าความถูกต้องในการติดตามหลายวัตถุ	36
3.6.2.5 ค่าความแม่นยำในการติดตามหลายวัตถุ	39
3.7 ค่าเฉลี่ยเรขาคณิต	39
บทที่ 4 งานวิจัยที่น่าสนใจ	40
4.1 ภาพรวมงานวิจัยที่น่าสนใจ	40
4.2 การตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน	41
4.2.1 ขั้นตอนการตรวจจับ	43
4.2.2 Latent SVM	45
4.3 การหาแตรีกเล็ดโดยวิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด	46
4.4 การปรับปรุงแตรีกเล็ดด้วยการประมาณตำแหน่ง	49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
4.4.1 การกำจัด Outlier	50
4.4.2 การปรับปรุงแทร็คเล็ท	54
4.5 การรวมแทร็คเล็ทเป็นเส้นทางการเดินและปรับปรุงความถูกต้อง ด้วยเทคนิคการปรับเรียบข้อมูล	54
4.5.1 การรวมแทร็คเล็ทเป็นเส้นทางการเดินโดยใช้การพิจารณาตาม เกณฑ์ของคลิกรกราฟที่สั้นที่สุด	58
4.5.2 แบ่งส่วนเส้นทางการเดินย่อยด้วยระยะห่างระหว่างแทร็คเล็ท	58
4.5.3 ประมาณเส้นทางการเดินด้วยเทคนิคการปรับเรียบ	61
4.5.4 กำจัด Outlier ของเส้นทางการเดินย่อย	63
4.5.5 คัดเลือกส่วนของเส้นทางการเดินสมบูรณ์ของแต่ละคน	63
บทที่ 5 การทดลองและผลการทดลอง	66
5.1 ชุดข้อมูล	66
5.2 การวิเคราะห์ค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมสำหรับการจัดค่าที่ไม่มากที่สุด เพื่อเพิ่มประสิทธิภาพการตรวจจับบุคคลจากภาพนิ่ง	67
5.3 การวิเคราะห์ค่าขีดแบ่งสำหรับการหาค่าที่อยู่ในเกณฑ์และช่วงขอบเขตข้อมูลที่ใช้ ในการประมาณเส้นทางการเดิน	70
5.4 การวัดประสิทธิภาพของการติดตามวัตถุเทียบกับวิธีการ GMCP-Tracker	73
บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ	75
6.1 การวิเคราะห์และสรุปผลการดำเนินงานวิจัย	75
6.1.1 การวิเคราะห์และสรุปผลค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมสำหรับการจัดค่าที่ไม่ มากที่สุด เพื่อเพิ่มประสิทธิภาพการตรวจจับบุคคลจากภาพนิ่ง	75
6.1.2 การวิเคราะห์และสรุปค่าขีดแบ่งสำหรับการหาค่าที่อยู่ในเกณฑ์และช่วงขอบเขตข้อมูล ที่ใช้ในการประมาณเส้นทางการเดิน	75
6.1.3 การวิเคราะห์และสรุปการวัดประสิทธิภาพของการติดตามวัตถุเทียบกับ วิธีการ GMCP-Tracker	76
6.2 ข้อจำกัดและขอบเขตของงานวิจัย	76
6.3 แนวทางในการพัฒนา	76
เอกสารอ้างอิง	78
ภาคผนวก ผลงานที่ได้รับการตีพิมพ์	80
ประวัติผู้เขียน	90

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงการเปรียบเทียบอัลกอริทึมจากงานวิจัยอ้างอิง [4] [2] กับงานวิจัยที่นำเสนอ.....	10
3.1 สรุปผลบวกจริง, ผลบวกกลาง, ผลลบจริง และ ผลลบกลาง.....	31
5.1 ค่าความสมบูรณ์ของการติดตามของเส้นทางการเดินที่เลือกใช้ วิธีการปรับเรียบข้อมูล Robust Local Regression และ Savitzky-Golay	72
5.2 ค่าความสมบูรณ์ของการติดตามของเส้นทางการเดินที่ได้จากการปรับค่าขอบเขตช่วงข้อมูลที่ใช้ ในการประมาณเส้นทางการเดินและค่าขีดแบ่งระยะห่างในการเลือกค่าที่อยู่ในเกณฑ์	72
5.3 การเปรียบเทียบประสิทธิภาพของงานวิจัยที่นำเสนอกับวิธี GMCP[2].....	73



สารบัญรูป

รูปที่		หน้า
2.1	กราฟสองฝ่ายและการจับคู่.....	6
2.2	ภาพรวมงานวิจัย[4].....	7
2.3	การจับคู่สองฝ่ายของกราฟและการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด	8
2.4	แผนภาพแสดงขั้นตอนของงานวิจัย [2]	9
3.1	การแบ่งเซลล์ บล็อก ของภาพที่จะนำมาหาคุณลักษณะ	11
3.2	ค่าความถี่ของทิศเกรเดียนท์ภายในเซลล์	13
3.3	ค่าคุณลักษณะความถี่ของทิศเกรเดียนท์ในแต่ละบล็อก	14
3.4	แสดงการรวมค่าเวกเตอร์คุณลักษณะ	14
3.5	ขั้นตอนการทำการจัดค่าที่ไม่มากที่สุดสำหรับการตรวจจับวัตถุ.....	15
3.6	แสดงการหาแตร็คเส้นจาก 6 เฟรมภายใน 1 ส่วนย่อยของวิดีโอ	19
3.7	คุณลักษณะการเคลื่อนไหวเพื่อการหาแตร็คเส้น.....	21
3.8	การรวมแตร็คเส้นกลายเป็นเส้นทางการเดินในกรณีต่างๆ.....	24
3.9	การเชื่อมโยงแตร็คเส้นเป็นเส้นทางการเดิน	26
3.10	การปรับเรียบข้อมูลด้วยวิธี lowess	28
3.11	การคำนวณค่าส่วนเหลือ	29
3.12	ขั้นตอนการปรับเรียบข้อมูลด้วยสมการถดถอยโดยพิจารณาเฉพาะส่วนแบบแกร่ง	30
3.13	ข้อมูลตำแหน่งจริงและ ผลลัพธ์การตรวจจับที่เป็น ผลบวกจริง ผลบวกลวง ผลลบลวงและผลลบจริง.....	31
3.14	การแยกแยะ TP,FP,FN ในกรณีต่างๆ โดยในตัวอย่างนี้กำหนดค่า $T_{ov} = 0.18$ เป็นเกณฑ์สำหรับการแยกแยะ.....	32
3.15	การนับจำนวนวัตถุที่ติดตามพลาด จำนวนผลบวกลวง จำนวนของการติดตามตัวสลับบุคคล ในกรณีต่างๆ.....	38
4.1	ภาพรวมงานวิจัยที่นำเสนอ.....	40
4.2	โมเดลประเภทคน.....	42
4.3	อิมเมจพีระมิตและคุณลักษณะพีระมิต	42
4.4	ภาพรวมขั้นตอนการตรวจจับของงานวิจัย[1]	43
4.5	ขั้นตอนการปรับปรุงแตร็คเส้น.....	50
4.6	การทำ inlier จาก tracklet บนแกนต่างๆ.....	51
4.7	การทำ inlier บน tracklet ด้วยวิธีการ linear regression แบบ least squares ในแกน x-y.....	52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่	หน้า
4.8	เส้นตรงตัวแทนแทร็คเล็ทที่ทำนายจาก inlier บนแกนต่างๆ.....53
4.9	แทร็คเล็ทที่ปรับปรุงแล้วและเส้นตรงตัวแทนแทร็คเล็ท.....54
4.10	ผลเส้นทางการเดินที่มีการติดตามสลับตัวบุคคล..... 55
4.11	ขั้นตอนการรวมแทร็คเล็ทเป็นเส้นทางการเดิน และปรับปรุงความถูกต้องด้วยเทคนิค smooth data.....57
4.12	กรณีการเลือกคำตอบของเส้นทางการเดินสลับตัวบุคคล59
4.13	แสดงระยะห่างระหว่างแทร็คเล็ท..... 60
4.14	การแบ่งส่วนย่อยของเส้นทางการเดิน..... 60
4.15	การทำ smooth data บนแกนต่างๆ และเส้นประมาณเส้นทางการเดิน.....62
4.16	เส้นประมาณเส้นทางการเดินกับแทร็คเล็ทภายในคำตอบ เส้นทางการเดินขั้นต้นที่เป็นไปได้.....62
4.17	กราฟแสดงขอบเขตในการแบ่งแยก inlier และ outlier.....63
4.18	ส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% (สีเขียว) และส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier น้อยกว่า 50% (สีแดง).....64
4.19	คัดเฉพาะส่วนส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% เพื่อมาเป็นคำตอบเส้นทางการเดินของคน 1 คน.....64
4.20	พิกัดที่เป็น inlier (สีเขียว) และ outlier (สีแดง).....65
4.21	เส้นทางการเดินของคน 1 คน.....65
5.1	ชุดการทดลอง.....66
5.2	การทำการจัดค่าที่ไม่มากที่สุดโดยการใช้ค่าขีดแบ่งพื้นที่ทับซ้อนต่างๆ.....67
5.3	กราฟแสดงความสัมพันธ์ระหว่างค่าขีดแบ่งพื้นที่ทับซ้อน กับ TP FP และ FN.....68
5.4	กราฟแสดงความสัมพันธ์ระหว่างค่าขีดแบ่งพื้นที่ทับซ้อนกับ N-MODA N-MODP และ Geometric Mean.....69
5.5	ผลลัพธ์เส้นทางการเดินเทียบกับเส้นทางการเดินที่ถูกต้อง..... 71

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การติดตามหลากหลายวัตถุ (Multiple Object Tracking) คือ การระบุตำแหน่งหลายวัตถุ เพื่อสร้างเส้นทางการเดินของวัตถุและแยกแยะวัตถุ โดยพิจารณาจากตำแหน่งวัตถุในแต่ละเฟรมของวิดีโอ ซึ่งการติดตามหลากหลายวัตถุได้ถูกนำไปประยุกต์ในหลายด้าน เช่น วิดีโอเพื่อรักษาความปลอดภัย การวิเคราะห์กิจกรรม การปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ การควบคุมการจราจรและวิทยาการหุ่นยนต์ การรู้จำการเคลื่อนไหว การสร้างดัชนีภาพและการค้นคืนภาพ ระบบนาทางในยานพาหนะ เป็นต้น ดังนั้นการติดตามวัตถุจึงเป็นงานวิจัยที่สำคัญยิ่งในวงการของคอมพิวเตอร์วิทัศน์

ที่ผ่านมาได้มีหลากหลายงานวิจัยที่ศึกษาในเรื่องของการติดตามวัตถุ ปัญหาที่พบจากวิธีการต่างๆ เช่น การติดตามมีการติดตามที่ระบุสลับเป้าหมายกัน อันเนื่องมาจากวัตถุถูกบดบัง หรือในสถานการณ์ที่เต็มไปด้วยฝูงชน เป็นต้น เพื่อเพิ่มประสิทธิภาพการติดตามวัตถุ งานวิจัยที่นำเสนอได้นานาเทคนิคการปรับเรียบข้อมูลและการวิเคราะห์การถดถอยเชิงเส้นมาประยุกต์ใช้ร่วมกับวิธีการติดตามหลากหลายวัตถุที่พิจารณาข้อมูลตลอดช่วงระยะเวลาของวิดีโอ เพื่อปรับปรุงประสิทธิภาพผลลัพธ์เส้นทางการเดินที่ได้จากการติดตามวัตถุ และเพื่อลดปัญหาการติดตามระบุสลับเป้าหมายกัน

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

- 1.2.1 เพื่อสามารถศึกษา และทำความเข้าใจงานวิจัยของการติดตามวัตถุและตรวจจับวัตถุ
- 1.2.2 เพื่อให้มีความรู้ ความเข้าใจ ในทฤษฎีพื้นฐานของการตรวจจับวัตถุ การติดตามวัตถุ และทฤษฎีที่เกี่ยวข้อง
- 1.2.3 เพื่อศึกษา ออกแบบ และประยุกต์ใช้การปรับเรียบข้อมูลและวิธีการถดถอยเชิงเส้นในการเพิ่มประสิทธิภาพของผลลัพธ์เส้นทางการเดินที่ได้จากการติดตามวัตถุ
- 1.2.4 เพื่อสามารถประเมินประสิทธิภาพของวิธีการตรวจจับวัตถุและติดตามวัตถุ รวมทั้งวิเคราะห์และสรุปผลที่ได้จากงานวิจัย
- 1.2.5 เพื่อเป็นแหล่งข้อมูลอ้างอิงสำหรับผู้ที่สนใจศึกษางานวิจัยด้านนี้ต่อไป

1.3 สมมติฐานของการศึกษา

งานวิจัยที่นำเสนอได้ทำการศึกษาด้วยการติดตามรูปคน ดังนั้นลักษณะของการติดตามภายในวิดีโอ จะเป็นลักษณะของชุดข้อมูลที่เป็นรูปคนเดิน และผลลัพธ์ของการติดตามจะเป็นเส้นทางการเดินของบุคคลแต่ละคน โดยงานวิจัยที่นำเสนอจะเสนอ การนำผลลัพธ์ของการตรวจจับคนมาทำการเชื่อมโยงข้อมูลเพื่อหาเส้นทางการเดินในระยะสั้นๆ ก่อน ซึ่งในงานวิจัยนี้เรียกว่าแทร็คเล็ต (Tracklet)

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยแต่ละแทร็คเล็ทนั้นจะเป็นผลลัพธ์เส้นทางการเดินช่วงสั้นๆที่ได้จากการเชื่อมโยงข้อมูลภายในช่วงเฟรม โดยช่วงเฟรมที่แบ่งจะแบ่งจ านวนเฟรมเป็นช่วงสั้นๆ ภายใต้สมมุติฐานที่ว่า เป็นช่วงที่คนเดินด้วยความเร็วคงที่และเดินเป็นวิถีเส้นตรง ต่อมางานวิจัยที่นำเสนอนี้จะปรับปรุงประสิทธิภาพแทร็คเล็ทด้วยการประมาณตำแหน่งจากการวิเคราะห์การถดถอยเชิงเส้น แล้วจึงนำแทร็คเล็ทไปรวมเป็นเส้นทางการเดิน (Trajectory) ที่สมบูรณ์ตลอดทั้งวิดีโอ ในขั้นตอนของการเชื่อมโยงข้อมูลเพื่อรวมแทร็คเล็ทเป็นผลลัพธ์เส้นทางการเดิน งานวิจัยที่นำเสนอใช้วิธีการปรับเรียบข้อมูลมาปรับปรุงประสิทธิภาพของการติดตามบุคคลและเพื่อลดปัญหาการติดตามระบุสลับตัวบุคคล

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย

วิธีการติดตามวัตถุที่นำเสนอในงานวิจัยนี้ อาศัยทฤษฎีและหลักการดังต่อไปนี้

- 1.4.1 การตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน[1] เป็นวิธีการที่ใช้ค่าความถี่ของทิศกระเดียนท์ มาพัฒนาต่อด้วยแนวคิดการฝึกสอนแบบแยกส่วนของวัตถุ
- 1.4.2 การติดตามวัตถุใช้วิธีการพิจารณาตามเกณฑ์ของคลัสกราฟที่สั้นที่สุดในการเชื่อมโยงข้อมูล[2]
- 1.4.3 การปรับปรุงผลลัพธ์เส้นทางการเดินระยะสั้น ด้วยวิธีการถดถอยเชิงเส้น (Linear regression)
- 1.4.4 การสร้างผลลัพธ์เส้นทางการเดินและปรับปรุงความถูกต้องด้วยเทคนิคการปรับเรียบข้อมูล (Smoothing Data)

1.5 ขอบเขตงานวิจัย

งานวิจัยที่ได้ทำ การศึกษานี้ มีขอบเขตการวิจัยดังต่อไปนี้

- 1.5.1 งานวิจัยนี้ได้ทดลองและท ำการศึกษาด้วยการติดตามรูปคน ดังนั้นผลลัพธ์ของการติดตามจะเป็น เส้นทางการเดินของบุคคลแต่ละคน
- 1.5.2 งานวิจัยนี้ท ำการแบ่งหาผลลัพธ์เส้นทางการเดินช่วงสั้นๆ จากช่วงส่วนย่อยวิดีโอก่อน โดยเรียกผลลัพธ์เส้นทางการเดินช่วงสั้นๆนี้ว่าแทร็คเล็ท (Tracklet) โดยแทร็คเล็ทของงานวิจัยนี้มีความหมายว่าเป็นผลลัพธ์เส้นทางการเดินระยะสั้นๆที่ได้จากการเชื่อมโยงข้อมูลวิดีโอภายในช่วงเฟรมระยะสั้นๆที่คนเดินด้วยความเร็วคงที่และเดินเป็นวิถีเส้นตรง หลังจากนั้นจึงค่อยท ำการเชื่อมโยงข้อมูลในการรวมแทร็คเล็ทนี้ให้กลายเป็นผลลัพธ์เส้นทางการเดินตลอดทั้งวิดีโอ
- 1.5.3 ชุดการทดลองของงานวิจัยนี้ เป็นการทดลองบนภาพนิ่งจากแต่ละเฟรมวิดีโอที่เรียงลำดับกัน

1.6 ขั้นตอนการวิจัย

- 1.6.1 กำหนดวัตถุประสงค์และขอบเขตของงานวิจัย
- 1.6.2 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการติดตามวัตถุ
- 1.6.3 วิเคราะห์และเปรียบเทียบจุดเด่น จุดด้อยของวิธีการติดตามวัตถุจากงานวิจัยที่ศึกษา เพื่อตั้งสมมติฐานและกำหนดแนวทางในการวิจัย
- 1.6.4 เตรียมฐานข้อมูลวิดีโอ ส าหรับตรวจสอบสมมติฐานและแนวความคิดของงานวิจัย รวมทั้งวิธีการตรวจสอบประสิทธิภาพของงานวิจัย
- 1.6.5 ออกแบบและพัฒนาระบบการติดตามวัตถุ ตามแนวความคิดและสมมติฐานของงานวิจัย
- 1.6.6 ทดสอบ และวิเคราะห์ประสิทธิภาพของระบบการติดตามวัตถุที่น าเสนอ
- 1.6.7 สรุปผลและน าเสนอผลงานวิจัย

1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย

เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย มีดังต่อไปนี้

- 1.7.1 เครื่องคอมพิวเตอร์ส่วนบุคคลประกอบด้วยหน่วยประมวลผลกลาง Intel Core i5 และหน่วยความจำ (RAM) 8 GB จำนวน 1 เครื่อง
- 1.7.2 ระบบปฏิบัติการ Windows 10
- 1.7.3 โปรแกรม Matlab รุ่น R2015a

1.8 โครงสร้างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้แบ่งออกเป็น 6 บท แต่ละบทประกอบด้วยเนื้อหาดังต่อไปนี้

- บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา ความมุ่งหมายและวัตถุประสงค์ของการศึกษา สมมติฐานของการศึกษา ทฤษฎีหรือแนวคิดที่ใช้ในงานวิจัย ขอบเขตงานวิจัย ขั้นตอนการวิจัย และเครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย
- บทที่ 2 กล่าวถึงงานวิจัยที่เกี่ยวข้องกับการติดตามหลากหลายวัตถุ ภาพรวมแต่ละงานวิจัยที่เกี่ยวข้อง รวมทั้งข้อดีและข้อจ ากัดของแต่ละวิธี
- บทที่ 3 กล่าวถึงทฤษฎีพื้นฐานที่เกี่ยวข้องกับงานวิจัยได้แก่ ค่าความถี่ของทิศเกรเดียนท์ (Histogram of Oriented Gradient :HOG) การขจัดค่าที่ไม่มากที่สุด (Nonmaximum suppression) สำหรับการตรวจจับวัตถุ การติดตามหลากหลายวัตถุโดยใช้เทคนิคการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด (GMCP-Tracker : Global Multi-object Tracking Using Generalized Minimum Clique Graphs) การถดถอยเชิงเส้น (Linear Regression) การปรับเรียบข้อมูลแบบถดถอยโดยพิจารณาเฉพาะส่วน (Local

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Regression Smoothing) การวัดประสิทธิภาพส าหรับการตรวจจับวัตถุและการติดตามวัตถุ

บทที่ 4 กล่าวถึงภาพรวมและขั้นตอนค านินการในงานวิจัยนี้ โดยมีขั้นตอนการตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน การหาแพริคเล็ทโดยวิธีการพิจารณาตามเกณฑ์ของคลิคกราฟที่สั้นที่สุด การปรับปรุงแพริคเล็ทด้วยการประมาณตาแหน่ง และการรวมแพริคเล็ทเป็นเส้นทางการเดินและปรับปรุงความถูกต้องด้วยเทคนิคการปรับเรียบข้อมูล

บทที่ 5 การทดลอง ผลการทดลอง และการวิเคราะห์ผล

บทที่ 6 สรุปผลการทดลอง ข้อเสนอแนะ และแนวทางในการพัฒนาต่อ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

งานวิจัยที่เกี่ยวข้อง

การติดตามหลากหลายวัตถุ (Multiple Object Tracking) คือ การระบุตำแหน่งหลายวัตถุ เพื่อสร้างเส้นทางการเดินของวัตถุและแยกแยะวัตถุ โดยพิจารณาจากตำแหน่งวัตถุในแต่ละเฟรมของวิดีโอ ซึ่งการติดตามหลากหลายวัตถุได้ถูกนำไปประยุกต์ในหลายด้าน เช่น วิดีโอเพื่อรักษาความปลอดภัย การวิเคราะห์กิจกรรม การปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์ การควบคุมการจราจรและวิทยาการหุ่นยนต์ การรู้จำการเคลื่อนไหว การสร้างดัชนีภาพและการค้นคืนภาพ ระบบนาทางในยานพาหนะ เป็นต้น การติดตามวัตถุจึงเป็นงานวิจัยที่สำคัญยิ่งในวงการของคอมพิวเตอร์วิทัศน์

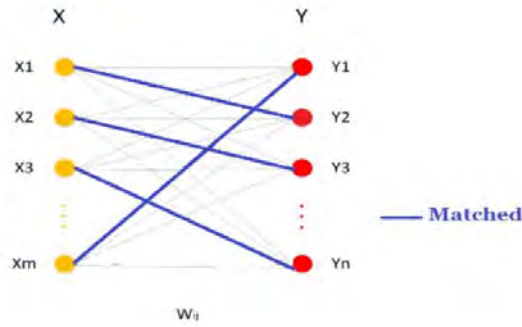
วิธีการติดตามวัตถุจากผลการตรวจจับ (tracking-by-detection) เป็นวิธีที่นิยมในการทำการติดตามวัตถุ โดยวิธีการนี้มีสองขั้นตอนหลัก ได้แก่ การตรวจจับวัตถุ (Object Detection) และการเชื่อมโยงข้อมูล (Data Association) สำหรับขั้นตอนการตรวจจับวัตถุ ตัวตรวจจับวัตถุที่ผ่านการฝึกสอนจะทำการหาตำแหน่งวัตถุที่ต้องการตรวจจับจากเฟรมวิดีโอแต่ละเฟรม เมื่อวัตถุที่อยู่ในภาพถูกตรวจจับ ผลลัพธ์การตรวจจับจะเป็นตัวแทน เพื่อนำไปสู่ขั้นตอนของการติดตามวัตถุ โดยในขั้นตอนนี้จะเชื่อมโยงระหว่างผลการตรวจจับวัตถุขึ้นเดียวกันที่อยู่ต่างเฟรมกัน โดยวิธีการติดตามวัตถุที่เป็นที่นิยมคือวิธีการเชื่อมโยงข้อมูล

วิธีการเชื่อมโยงข้อมูล (Data Association Method) ได้กำหนดปัญหาการติดตามวัตถุ ว่าเป็น การเลือกและการแบ่งกลุ่มของผลการตรวจจับวัตถุโดยพิจารณาจากช่วงเวลาทั้งหมด โดยสามารถแบ่งวิธีการเชื่อมโยงข้อมูลเป็นสองประเภท ได้แก่ การเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วน (Local optimization based data association) และการเชื่อมโยงข้อมูลโดยพิจารณาจากภาพรวม (Global optimization based data association) [3]

2.1 การเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วน

(Local optimization based data association)

การเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วนจะเชื่อมโยงข้อมูลโดยพิจารณาจากช่วงเฟรมวิดีโอแค่เพียงไม่กี่เฟรม โดยส่วนมากจะพิจารณาจากแค่เพียง 2 เฟรมหรือมากกว่านั้นเพียงเล็กน้อย วิธีการที่เป็นที่นิยมของการเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วนคือ การจับคู่สองฝ่าย (Bipartite Matching) โดยอัลกอริทึมที่นิยมใช้ในการแก้ปัญหาการจับคู่สองฝ่ายคือ อัลกอริทึมฮังการี (Hungarian algorithm) ซึ่งใช้กันอย่างแพร่หลายในการคำนวณความสัมพันธ์ระหว่างแทร็คเล็ต (Tracklet) (แทร็คเล็ต หมายถึงเส้นทางการเดินระยะสั้น) และผลลัพธ์การตรวจจับวัตถุ



รูปที่ 2.1 กราฟสองฝ่ายและการจับคู่

งานวิจัย[4] ใช้วิธีการจับคู่กราฟสองฝ่าย (bipartite graph matching) สำหรับการติดตามเป้าหมายโดยที่โหนดของกราฟสองฝ่ายจะสอดคล้องกับเป้าหมายที่อยู่ในระหว่างสองเฟรมที่ติดกัน และค่าน้ำหนักจะพิจารณาจากความคล้ายคลึงระหว่างเป้าหมายที่อยู่คนละเฟรม โดยค่าน้ำหนักจะพิจารณาจากการพิจารณาระยะห่างระหว่างตำแหน่ง ขนาดของผลการตรวจจับ ค่าความแตกต่างของค่าฮิสโตแกรมสี HSV และ ค่าฮิสโตแกรม CS-BLP โดยกำหนดให้แต่ละโหนดผลการตรวจจับคือ $N_i = \{p_i, s_i, f_i\}$ โดย p_i คือตำแหน่ง s_i คือขนาด และ f_i คือคุณลักษณะสี

$$S_{ij}(N_i, N_j) = S_{\text{pos}}(N_i, N_j)S_{\text{size}}(N_i, N_j)S_{\text{feature}}(N_i, N_j) \quad (2.1)$$

$$S_{\text{pos}}(N_i, N_j) = \gamma_{\text{pos}} \exp \left[-\frac{(x_i - x_j)^2}{\sigma_x^2} \right] \exp \left[-\frac{(y_i - y_j)^2}{\sigma_y^2} \right] \quad (2.2)$$

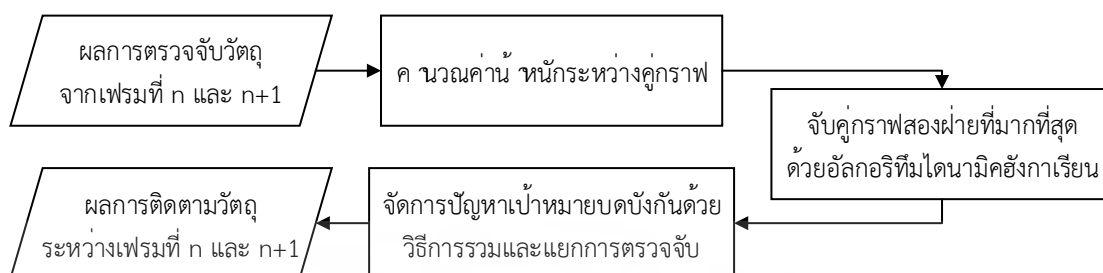
$$S_{\text{size}}(N_i, N_j) = \gamma_{\text{size}} \exp \left[-\frac{(s_i - s_j)^2}{\sigma_s^2} \right] \quad (2.3)$$

$$S_{\text{feature}}(N_i, N_j) = B_{\text{HSV}}(N_i, N_j)B_{\text{CS-LBP}}(N_i, N_j) \quad (2.4)$$

จากสมการที่ 2.1 กำหนดให้ $S_{ij}(N_i, N_j)$ คือค่าน้ำหนักระหว่างโหนด จากสมการที่ 2.2 $S_{\text{pos}}(N_i, N_j)$ คือค่าคุณลักษณะระยะห่างระหว่างโหนด สมการที่ 2.3 $S_{\text{size}}(N_i, N_j)$ คือค่าคุณลักษณะขนาดของผลการตรวจจับ สมการที่ 2.4 $S_{\text{feature}}(N_i, N_j)$ คือค่าคุณลักษณะความแตกต่างของค่าฮิสโตแกรมสี HSV และ CS-BLP โดยที่ $B(\cdot)$ คือค่าระยะห่างแบบ Bhattacharyya ระหว่างสองฮิสโตแกรม γ_{pos} และ γ_{size} คือค่านอร์มัลไลซ์

เมื่อได้ค่าน้ำหนักระหว่างคู่กราฟ การหาความสอดคล้องกันระหว่างเป้าหมายทั้งสองเฟรมจะถูกคำนวณในรูปแบบปัญหาการจับคู่ค่าที่มากที่สุดซึ่งแก้ด้วยอัลกอริทึมไดนามิกฮังการี (The Dynamic Hungarian algorithm) ต่อมาแก้ปัญหาวัดถูกบดบังด้วยขั้นตอนการรวมผลและแยกผล เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การตรวจจับ พื้นที่ที่มีโอกาสถูกบดบังจะถูกคัดคะแนนจากพื้นที่ทับซ้อนระหว่างเป้าหมายที่มีปัญหา การถูกบดบัง



รูปที่ 2.2 ภาพรวมงานวิจัย[4]

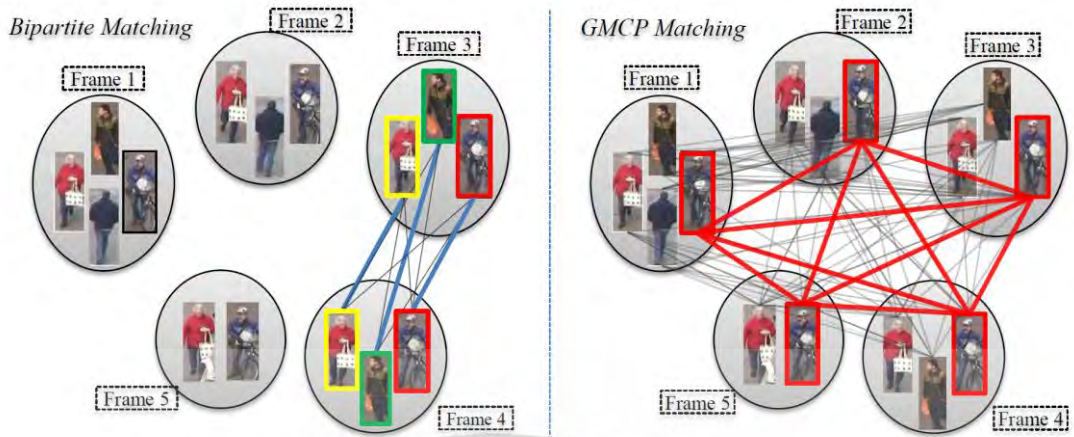
สำหรับข้อจำกัดของการเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วน ในกรณีที่คนเดินถูกบดบัง อาจ ทำให้การติดตามไปเลือกคำตอบคนอื่นที่มีลักษณะใกล้เคียงกัน เนื่องจากใช้ข้อมูลจากภาพเพียงไม่กี่ เฟรมมาพิจารณาในช่วงสั้นๆมาพิจารณา ทำให้เกิดการติดตามสลับตัวบุคคลเกิดขึ้นได้ง่าย และผลการ ติดตามที่ผิดพลาดจะส่งผลกระทบต่อเนื่องเป็นลูกโซ่ในการติดตามเฟรมถัดไปอีกด้วย

2.2 การเชื่อมโยงข้อมูลโดยพิจารณาจากภาพรวม

(Global optimization based data association)

การเชื่อมโยงข้อมูลโดยพิจารณาเฉพาะส่วน อาศัยพื้นฐานการเชื่อมโยงข้อมูลบุคคลในช่วงเฟรม สั้นๆ เช่นที่ละคู่เฟรม ทำให้สามารถเกิดปัญหาการระบุตัวบุคคลสลับกัน อันเนื่องมาจากเหตุการณ์ที่ บุคคลโดนบดบังเป็นระยะเวลาสั้นๆ และในสถานการณ์ที่เต็มไปด้วยฝูงชน ดังนั้นการเชื่อมโยงข้อมูล โดยพิจารณาจากภาพรวมจะสามารถแก้ไขปัญหาดังกล่าวได้ อันเนื่องมาจากการเชื่อมโยงข้อมูลโดย พิจารณาจากภาพรวมเป็นการเชื่อมโยงข้อมูลบุคคลจากทุกเฟรมตลอดทั้งวิดีโอพร้อมกัน

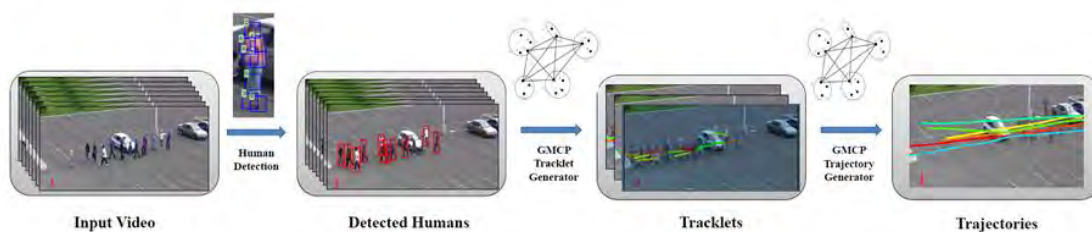
ดังนั้นเพื่อลดปัญหาการระบุตัวบุคคลผิดพลาด งานวิจัย [2] จึงได้นำเสนอวิธีการพิจารณาข้อมูล บุคคลตลอดช่วงระยะเวลาของวิดีโอ โดยเชื่อมโยงข้อมูลซึ่งใช้ลักษณะการเคลื่อนไหวและลักษณะ รูปร่างพื้นฐานของมนุษย์ โดยเลือกพิจารณาจากเหตุการณ์ทั้งหมด ซึ่งแตกต่างจากวิธีการเชื่อมโยง ข้อมูลโดยการพิจารณาตำแหน่งของวัตถุในเชิงเวลาที่พิจารณาจากภาพเพียงไม่กี่เฟรมซึ่งจะทำให้เกิด ปัญหาต่างๆ ที่ได้กล่าวมาแล้วข้างต้น โดยขั้นตอนการแก้ปัญหาเหล่านั้นจะทำการเชื่อมโยงวัตถุทีละหนึ่ง วัตถุรวมเข้าด้วยกันจากช่วงเวลาทั้งหมด โดยใช้วัตถุที่เหลือมาพิจารณาโดยอ้อม ในขั้นตอนการ แก้ปัญหานั้น งานวิจัย [2] ได้เลือกใช้เทคนิคการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด (Generalized Minimum Clique Graphs) มาใช้ในการเชื่อมโยงข้อมูล



รูปที่ 2.3 การจับคู่สองฝ่ายของกราฟและการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด

จากรูปที่ 2.3 เป็นภาพแสดงการจับคู่สองฝ่ายของกราฟ (ด้านซ้าย) และวิธีการจากงานวิจัย [2] (ด้านขวา) แผนผังนี้มี 6 เฟรม โดยเส้นสีเทาหมายถึงกราฟของข้อมูล ข่าที่มี การเชื่อมโยงข้อมูลบุคคลใน 1 รอบของอัลกอริทึม รูปด้านขวา เส้นสีแดงแสดงถึงกราฟย่อยที่เหมาะสมที่เป็นผลลัพธ์ของรอบนั้นๆ ซึ่งแสดงถึงกราฟย่อยของผลลัพธ์ที่ได้จากงานวิจัย [2] ซึ่งเป็นการแสดงการตรวจจับของบุคคล 1 คนตลอดระยะเวลาทั้งหมดที่พิจารณา ในทางตรงกันข้าม วิธีการจับคู่สองฝ่ายของกราฟจะไม่สนใจข้อมูลอื่นๆ ที่อยู่นอกพื้นที่ (ไม่คิดเฟรมอื่นนอกจากในเฟรมที่ 3 และ 4)

จากรูปที่ 2.4 แผนภาพของภาพรวมทั้งหมดในขั้นตอนการเชื่อมโยงข้อมูล ขั้นตอนแรกคือการตรวจจับบุคคลในแต่ละเฟรม ใช้วิธีการการตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน [1] ต่อมาทำการแบ่งข้อมูลวิดีโอเข้าออกเป็นส่วนย่อย (segment) ตามจำนวนที่กำหนด และทำการหาแทร็คเล็ตของการเดินเท้าในแต่ละส่วนย่อย โดยใช้วิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุดมาใช้ในการหาแทร็คเล็ต (Tracklet) ในขั้นตอนสุดท้าย หากการรวมแทร็คเล็ตที่พบในทุกส่วนย่อยของวิดีโอออกมาเป็นเส้นทางการเดินทั้งหมดของแต่ละคนในวิดีโอ สาเหตุเหตุการณ์ที่มีฝูงคนเดินบดบังกันเป็นช่วงระยะสั้นหรือนาน ผลการตรวจจับคนที่เกิดความผิดพลาด ซึ่งเหตุการณ์ดังกล่าวอาจส่งผลทำให้เกิดการติดตามบุคคลที่ผิดพลาดหรือเกิดการติดตามสลับบุคคลเกิดขึ้น ถ้าหากฝูงคนเดินบดบังกันเป็นช่วงระยะเวลาอันยาวนานอาจส่งผลทำให้บุคคลที่ถูกบดบังไม่มีผลลัพธ์แทร็คเล็ต งานวิจัย [2] ได้แก้ปัญหาดังกล่าวด้วยการเพิ่มโหมดสมมุติทั้งในระดับขั้นตอนของการหาแทร็คเล็ตและในระดับขั้นตอนของการสร้างเส้นทางการเดิน ซึ่งสาหรับรายละเอียดของงานวิจัย [2] จะขออธิบายเพิ่มเติมในบทที่ 3 ส่วนหัวข้อที่ 3.3



รูปที่ 2.4 แผนภาพแสดงขั้นตอนของงานวิจัย [2]

ในส่วนขั้นตอนการสร้างโหนดสมมุติในระดับการสร้างแพริคเล็ทของงานวิจัย [2] มีขั้นตอนของการทำนายตำแหน่งของโหนดสมมุติ โดยการสกัดค่า inlier ด้วยวิธีการ RANSAC ก่อนจะนำค่า inlier ไปประมาณวิถีเส้นตรงเพื่อทำนายตำแหน่งของโหนดสมมุติ พบว่าการแบ่งกลุ่มค่า inlier และค่า outlier ยังมีการพิจารณาที่ไม่มีประสิทธิภาพ ซึ่งส่งผลทำให้วิถีเส้นตรงที่ประมาณเกิดความผิดพลาดได้ และในขั้นตอนการสร้างโหนดสมมุติที่ใช้ทดแทนแพริคเล็ทของคนที่ไม่โดนบดบังเป็นระยะเวลาสั้น พบว่าค่าคุณลักษณะของโหนดสมมุติที่จำลองขึ้นยังส่งผลให้เกิดการเชื่อมโยงข้อมูลที่ผิดพลาด ส่งผลให้เกิดผลลัพธ์เส้นทางการเดินที่ไม่ถูกต้อง

งานวิจัยที่นำเสนอมีจุดมุ่งหมายเพื่อนำเสนอเทคนิคการปรับปรุงแพริคเล็ทให้มีประสิทธิภาพที่ดีมากยิ่งขึ้น และแก้ปัญหาการเลือกแพริคเล็ทที่ผิดพลาดเพื่อลดปัญหาการติดตามสลับบุคคล เพื่อให้ได้ผลลัพธ์เส้นทางการเดินที่ถูกต้องและเป็นคำตอบของบุคคลคนเดียวกันตลอดทั้งผลลัพธ์ โดยในตารางที่ 2.1 จะแสดงการเปรียบเทียบอัลกอริทึมจากงานวิจัยอ้างอิง [4] [2] กับงานวิจัยที่นำเสนอ

ตารางที่ 2.1 แสดงการเปรียบเทียบอัลกอริทึมจากงานวิจัยอ้างอิง [4] [2] กับงานวิจัยที่นำเสนอ

	งานวิจัยที่ [4]	งานวิจัยที่ [2]	งานวิจัยที่นำเสนอ
ประเภทของการเชื่อมโยงข้อมูล	Local	Global	Global
ค่าน้ำหนัก	<ul style="list-style-type: none"> - ระยะห่างระหว่างตำแหน่ง - ขนาดของผลการตรวจจับ - ค่าความแตกต่างของค่าฮิสโตแกรมสี HSV และ CS-BLP 	<ul style="list-style-type: none"> - ค่าคุณลักษณะการเคลื่อนที่ - ค่าคุณลักษณะรูปพรรณสัณฐาน 	<ul style="list-style-type: none"> - ค่าคุณลักษณะการเคลื่อนที่ - ค่าคุณลักษณะรูปพรรณสัณฐาน
วิธีการเชื่อมโยงข้อมูล	The Dynamic Hungarian Algorithm	Generalized Minimum Clique Graphs (GMCP)	<ul style="list-style-type: none"> - Generalized Minimum Clique Graphs (GMCP) - การปรับปรุงแทร็คเล็ตด้วยการประมาณตำแหน่ง - การปรับปรุงความถูกต้องของเส้นทางการเดินทาง ด้วยเทคนิคการปรับเรียงข้อมูล
ข้อจำกัด	<p>การติดตามสลับตัวบุคคลได้ง่าย และส่งผลต่อเนื่องไปยังการติดตามต่อไป</p>	<ul style="list-style-type: none"> - ยังคงมีการติดตามสลับตัวบุคคลเกิดขึ้นอยู่มาก - ตำแหน่งผลลัพธ์การติดตามมีความคลาดเคลื่อน 	<ul style="list-style-type: none"> - ลดปัญหาการติดตามสลับบุคคล - เพิ่มประสิทธิภาพผลลัพธ์การติดตามที่แม่นยำมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

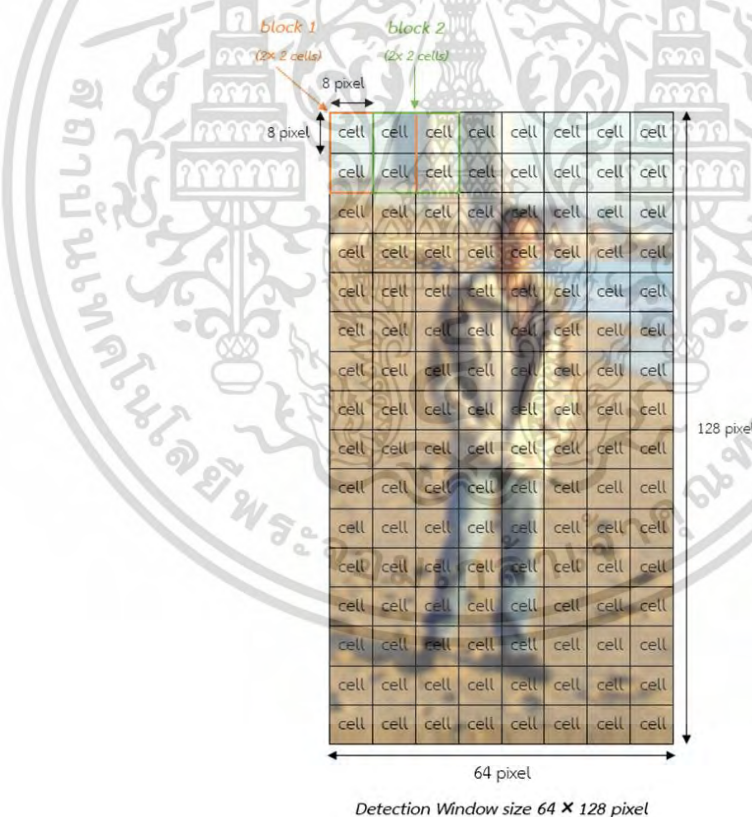
ทฤษฎีพื้นฐานที่เกี่ยวข้อง

3.1 ค่าความถี่ของทิศเกรเดียนท์ (Histogram of Oriented Gradient : HOG)

ค่าความถี่ของทิศเกรเดียนท์เป็นวิธีการพื้นฐานที่นิยมนามาใช้ในการสกัดคุณลักษณะของรูปคนออกมา เพื่อใช้ในการตรวจจับคน หรือจะเป็นวัตถุชนิดอื่นก็ได้

3.1.1 ตัวแปรเริ่มต้นในการสกัดคุณลักษณะค่าความถี่ของทิศเกรเดียนท์

หากการแบ่งภาพออกเป็นส่วนหลายส่วน กำหนดให้เรียกส่วนย่อยนี้ว่า เซลล์ (cells) โดยแต่ละเซลล์จะมีหน้าที่ในการเก็บค่าความถี่ของทิศเกรเดียนท์ภายในเซลล์ ซึ่งจำนวนทิศทางที่เก็บในแต่ละเซลล์นั้นเรียกว่า บิน (bins) และเมื่อนาฮิสโตแกรมจากแต่ละเซลล์มารวมกัน กำหนดให้เรียกกลุ่มที่ได้จากการรวมเซลล์นี้ว่า บล็อก (block)



รูปที่ 3.1 การแบ่งเซลล์ บล็อก ของภาพที่จะนำมาหาคุณลักษณะ

จากรูปที่ 3.1 เป็นตัวอย่างการกำหนดขนาดภาพที่จะนำมาคุณลักษณะสำหรับการตรวจจับ 1 คน เท่ากับ 64x128 พิกเซล กำหนดให้เซลล์มีขนาด 8x8 พิกเซล บล็อกมีเซลล์จำนวน 2x2 เซลล์ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 การคำนวณค่าเวกเตอร์ความลาดชัน (Gradient Vector) ค่าขนาดการเปลี่ยนแปลงของความเข้มแสง (Magnitude Gradient) และทิศทางของเกรเดียนท์เวกเตอร์

ในการคำนวณค่าเวกเตอร์ความลาดชัน ค่าขนาดการเปลี่ยนแปลงของความเข้มแสง และทิศทางของการเปลี่ยนแปลงของความเข้มแสง จะต้องใช้เทคนิคการหาอนุพันธ์อันดับหนึ่ง (First-Order Derivative) โดยการหาอนุพันธ์อันดับหนึ่ง จะนิยมใช้ในการหาตำแหน่งของการเปลี่ยนแปลงของความเข้มแสงในภาพซึ่งเป็นตำแหน่งขอบวัตถุ โดยในที่นี้จะใช้วิธีการสร้างเทมเพลตตัวกรอง แล้วนำค่าน้ำหนักของเทมเพลตตัวกรองไปทำคอนโวลูชัน (Convolution) กับภาพระดับเทา (Gray scale) โดยสมการที่ใช้ในการหาการเปลี่ยนแปลงในแนวแกน X ดังนี้

$$G_x = [-1 \ 0 \ 1] * I \quad (3.1)$$

สมการที่ใช้ในการหาการเปลี่ยนแปลงในแนวแกน Y มีดังนี้

$$G_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * I \quad (3.2)$$

เมื่อได้ค่าการเปลี่ยนแปลงทั้งในแกน X และแกน Y แล้ว ก็จะสามารถหาค่าเวกเตอร์ความลาดชันจากสมการ

$$\nabla F = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \quad (3.3)$$

โดย ∇F เป็นค่าเวกเตอร์ความลาดชัน (Gradient Vector) ซึ่งจากเวกเตอร์ความลาดชันจะสามารถหาไปหาค่าขนาดการเปลี่ยนแปลงของความเข้มแสง (Magnitude Gradient) และทิศทางของเกรเดียนท์เวกเตอร์ ได้จากสมการ 3.4 และ 3.5

$$\nabla f = \sqrt{G_x^2 + G_y^2} \quad (3.4)$$

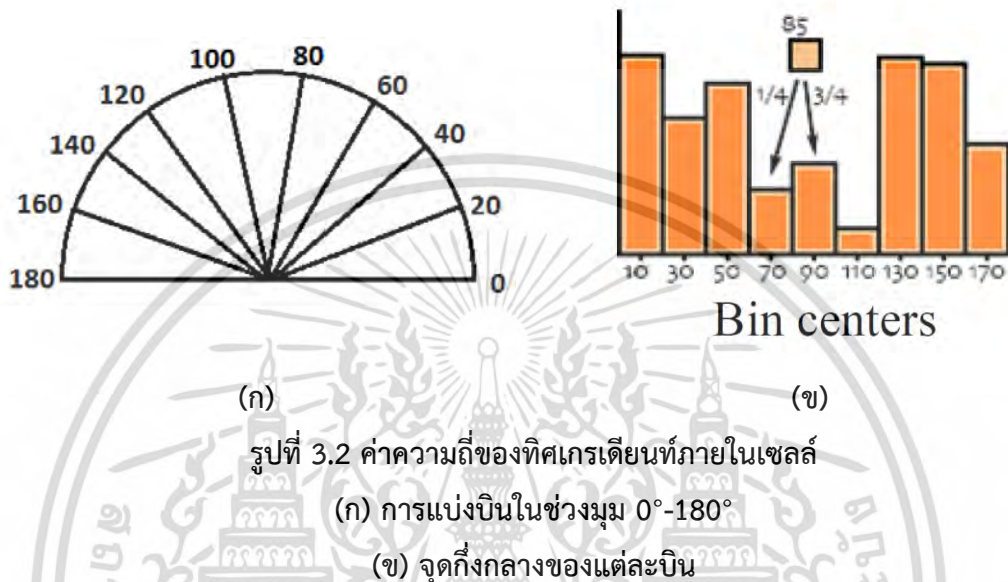
$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (3.5)$$

โดย ∇f เป็นค่าขนาดการเปลี่ยนแปลงของความเข้มแสงและ θ เป็นค่าทิศทางของเกรเดียนท์เวกเตอร์ ซึ่งวัดโดยเทียบกับแกน X

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 การคำนวณค่าความถี่ของทิศเกรเดียนท์ภายในเซลล์

การคำนวณค่าความถี่ของทิศเกรเดียนท์นั้น คือการนำค่าทิศทางของเกรเดียนท์เวกเตอร์ มาเป็นเกณฑ์ สำหรับการพิจารณาว่าจะนำค่า Magnitude Gradient ไปเพิ่มในค่าความถี่ช่วงไหนบ้าง และเพิ่มในแต่ละช่วงเป็นจำนวนเท่าไรบ้าง



จากรูปที่ 3.2 (ก) เป็นตัวอย่างการกำหนดให้มี 9 บินต่อฮิสโตแกรม ซึ่งในการแบ่งบินจะเป็นการแบ่งภายในระหว่างมุม 0°-180° ได้แก่มุม 0°-20° , 20°-40° , 40°-60° , 60°-80° , 80°-100° , 100°-120° , 120°-140° , 140°-160° , 160°-180°

จากรูปที่ 3.2 (ข) ถ้าค่าทิศทางของเกรเดียนท์เวกเตอร์เท่ากับ 85° จากรูป จะเห็นว่า 85° จะมีค่าอยู่ระหว่างจุดกึ่งกลางของชั้น 60°- 80° และ 80°- 100° ซึ่งมีระยะห่างจากจุดกึ่งกลางของชั้น 60°- 80° เท่ากับ 15° และจุดกึ่งกลางชั้น 80°- 100° เท่ากับ 5° ดังนั้นจากระยะห่างนี้จะได้อ่านน้ำหนักสำหรับการเพิ่มค่า Magnitude Gradient ในช่วง 60-80 เท่ากับ $\frac{1}{4}$ และสำหรับช่วง 80-100 เท่ากับ $\frac{3}{4}$ ดังนั้นจะเพิ่มค่าความถี่ของทิศเกรเดียนท์ในช่วง 60-80 เท่ากับ $\frac{1}{4} \times \nabla f$ และ $\frac{3}{4} \times \nabla f$ ซึ่งเมื่อเสร็จขั้นตอนนี้ก็เท่ากับว่าทุกเซลล์จะมีค่าความถี่ของทิศเกรเดียนท์แล้ว

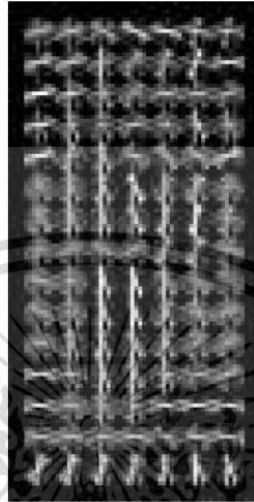
3.1.4 นอร์มัลไลเซชันบล็อก (Block Normalization)

นำค่าความถี่ของทิศเกรเดียนท์จากแต่ละเซลล์ที่อยู่ในบล็อก มาทำการนอร์มัลไลเซชันด้วยสมการที่ 3.6

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (3.6)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

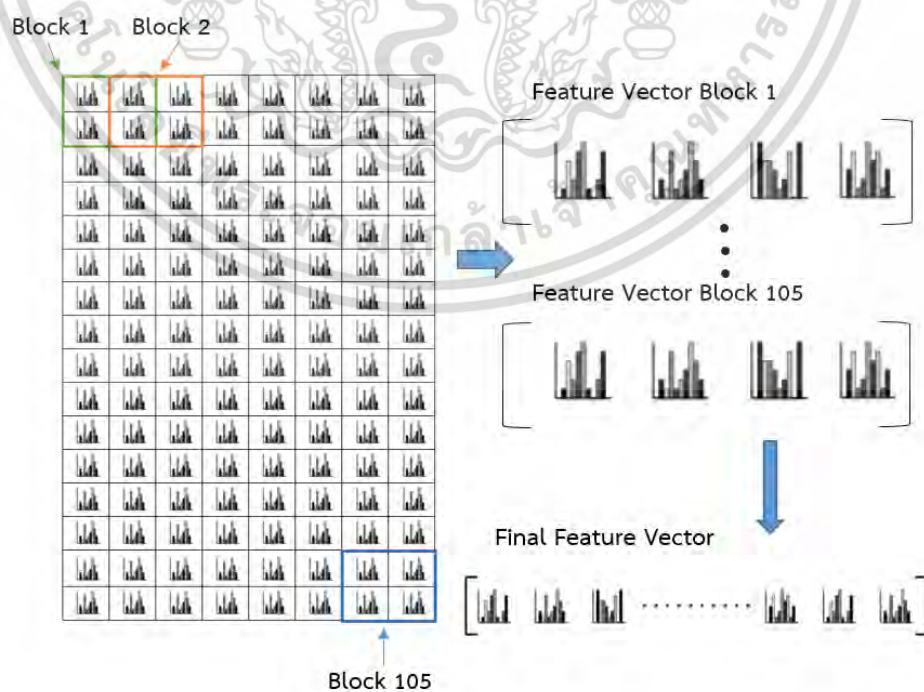
จากสมการที่ 3.6 f คือค่าเวกเตอร์คุณลักษณะในแต่ละบล็อก กำหนดให้ v เป็นเวกเตอร์ที่เก็บค่าความถี่ของทิศเกรเดียนท์ในแต่ละบล็อก $\|v\|_2$ คือ L2-norm และ e คือค่าคงที่จำนวนน้อยๆ เพื่อไม่ให้ส่วนเป็น 0 ในกรณีที่ค่าทุกค่า v เท่ากับ 0



รูปที่ 3.3 ค่าคุณลักษณะความถี่ของทิศเกรเดียนท์ในแต่ละบล็อก

3.1.5 ค่าเวกเตอร์คุณลักษณะ

จาก 3.1.4 แต่ละบล็อกจะมีค่าเวกเตอร์คุณลักษณะ นาเวกเตอร์ของแต่ละบล็อกมาเชื่อมต่อกันให้เป็นเวกเตอร์เดียวกันจะได้เป็นค่าเวกเตอร์คุณลักษณะของหน้าตาการตรวจจับนั้นๆ



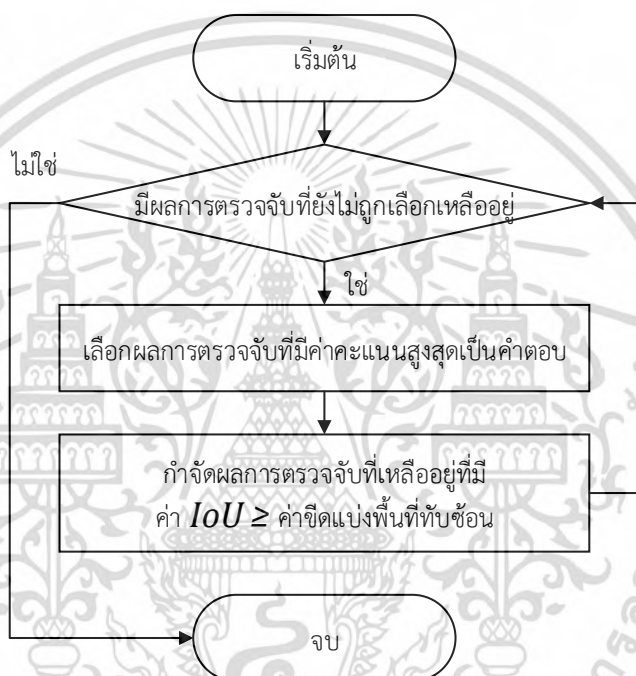
รูปที่ 3.4 แสดงการรวมค่าเวกเตอร์คุณลักษณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การขจัดค่าที่ไม่มากที่สุด (Nonmaximum suppression)

สำหรับการตรวจจับวัตถุ

หลังจากการทำการตรวจจับวัตถุเป็นปกติที่จะมีผลการตรวจจับที่ทับซ้อนกันหลายผลลัพธ์สำหรับวัตถุแต่ละชิ้น อันเนื่องมาจากสาเหตุที่ วิธีการตรวจจับวัตถุได้ทำการตรวจจับที่ภาพหลากหลายขนาดด้วยเทคนิคอิมเมจพีระมิด เป็นต้น วิธีการขจัดค่าที่ไม่มากที่สุดสำหรับการตรวจจับมีจุดมุ่งหมายคือ การลดผลการตรวจจับที่ซ้ำซ้อนที่เป็นค ทอบของวัตถุชิ้นเดียวกัน และเลือกให้เหลือแค่ผลการตรวจจับ 1 ค าตอบที่ดีที่สุดส ำหรับวัตถุ 1 ชิ้น



รูปที่ 3.5 ขั้นตอนการทำการขจัดค่าที่ไม่มากที่สุดสำหรับการตรวจจับวัตถุ

โดยขั้นตอนของการขจัดค่าที่ไม่มากที่สุดสำหรับการตรวจจับวัตถุ เริ่มต้นหลังจากที่วิธีการตรวจจับวัตถุคาดคะเนผลกรอบการตรวจจับวัตถุ โดยแต่ละผลการตรวจจับถูกนิยามได้จากกรอบสี่เหลี่ยมล้อมรอบวัตถุ (bounding box) และค่าคะแนนที่ได้จากวิธีการตรวจจับวัตถุ จากรูปที่ 3.5 จะมีขั้นตอนของการขจัดค่าที่ไม่มากที่สุด ส ำหรับการตรวจจับวัตถุดังต่อไปนี้

- 1) ถ้าหากยังมีกลุ่มผลการตรวจจับที่ยังไม่ถูกเลือกมาเป็นคำตอบเหลืออยู่จะทาขั้นตอนที่ 2 ต่อไป แต่ถ้าหากไม่มีผลการตรวจจับที่ยังไม่ถูกเลือกมาเป็นคำตอบเหลืออยู่แล้ว จะถือว่าสิ้นสุดการท างาน
- 2) ทาการเลือกผลการตรวจจับที่มีค่าคะแนนสูงสุดจากกลุ่มผลที่ยังไม่ถูกเลือก เพื่อมาเป็นค ำตอบ

- 3) ทาการคิดค่า IoU (Intersection over Union) จากสมการที่ 3.7 เพื่อหาพื้นที่ทับซ้อนระหว่างผลการตรวจจับที่ถูกเลือกในแต่ละรอบเทียบกับผลการตรวจจับที่ยังไม่ถูกเลือก

$$IoU = \frac{|D_1 \cap D_2|}{|D_1 \cup D_2|} \quad (3.7)$$

โดยกำหนดให้ D_1 คือผลการตรวจจับวัตถุที่ถูกเลือกในแต่ละรอบ และ D_2 คือผลการตรวจจับวัตถุที่ยังไม่ถูกเลือก

ถ้าหากพื้นที่ทับซ้อนของผลการตรวจจับที่ถูกเลือกในแต่ละรอบกับผลที่ยังไม่ถูกเลือก พื้นที่ทับซ้อนคู่ใดมีค่ามากกว่าหรือเท่ากับค่าขีดแบ่งพื้นที่ทับซ้อน มีโอกาสที่วัตถุสองชิ้นนั้นจะเป็นผลการตรวจจับของวัตถุชิ้นเดียวกัน จึงต้องทำการกำจัดผลการตรวจจับที่ยังไม่ถูกเลือกออกไป เพื่อลดความซ้ำซ้อน โดยจะท ำขั้นตอนนี้กับทุกผลการตรวจจับที่ยังไม่ถูกเลือกที่เหลือทั้งหมด

- 4) เมื่อจบขั้นตอนที่ 3 จะวนกลับไปท ำขั้นตอนที่ 1 ใหม่อีกรอบ

3.3 การติดตามหลากหลายวัตถุโดยใช้เทคนิคการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด (GMCP-Tracker : Global Multi-object Tracking Using Generalized Minimum Clique Graphs) [2]

งานวิจัย [2] น ำเสนอวิธีการเชื่อมโยงข้อมูลซึ่งเป็นการรวมลักษณะรูปพรรณสัณฐานและลักษณะการเคลื่อนไหวเข้าด้วยกัน โดยนำลำดับตลอดทั้งเหตุการณ์มาใช้ในการเชื่อมโยงข้อมูล แต่จะเลือกพิจารณาวัตถุที่ละชิ้นแทนที่จะพิจารณาหลายๆชิ้นพร้อมกัน ซึ่งนี่เป็นการหลีกเลี่ยงกับปัญหาที่มีความซับซ้อน เช่น ปัญหาที่คนเดินถูกบดบังเป็นช่วงเวลานาน และการระบุตัวบุคคลสลับคนกัน ถึงแม้ว่าในการแก้ปัญหาจะเลือกพิจารณาทีละหนึ่งชิ้น แต่ก็ได้ใช้วัตถุชิ้นอื่นเป็นเกณฑ์ในการพิจารณาโดยอ้อมไปด้วย ทำให้มีข้อจำกัดน้อยกว่าวิธีการอื่นๆ อย่างเช่น วิธีการจำกัดการพิจารณาขอบเขตพื้นที่ในเชิงเวลา (limited-temporal-locality) เพราะวิธีการนี้มุ่งเน้นพิจารณาแต่เพียงข้อมูลในช่วงระยะเวลาที่ใกล้เคียงกัน ในขณะที่งานวิจัย [2] จะเป็นการพิจารณาจากวัตถุทั้งหมด

จากรูปที่ 2.4 ในบทที่ 2 แผนภาพของภาพรวมทั้งหมดในขั้นตอนการเชื่อมโยงข้อมูล ขั้นตอนแรกคือการตรวจจับบุคคลในแต่ละเฟรม ใช้วิธีการการตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน [1] ต่อมาได้ทำการแบ่งข้อมูลวิดีโอ ำออกเป็นส่วนย่อย (segment) ตามจ านวนที่กำหนด และทาการหาแตร็คเส้นของการเดินเท้าในแต่ละส่วนย่อย โดยใช้วิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุดมาใช้ในการหาแตร็คเส้น และขั้นตอนสุดท้าย ท ำการรวมแตร็คเส้นที่พบในทุกส่วนย่อยของวิดีโอออกมาเป็นเส้นทางการเดินทั้งหมดของแต่ละคนในวิดีโอ

แม้รูปพรรณสัณฐานของคนเดินเท้าจะมีลักษณะที่คล้ายกันตลอดทั้งวิดีโอ แต่รูปแบบของการเคลื่อนไหวมีแนวโน้มที่จะแตกต่างกันในระยะสั้นและระยะยาว ด้วยเหตุนี้ มันจึงเป็นการยากที่จะเอกสารนี้เป็นเอกสารที่สว่นไวสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไมออนุญาตให้นำไปเผยแพร่บนนิตยสารใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำลองการเคลื่อนไหวของคนหนึ่งคนในช่วงระยะเวลาสั้นๆ โดยไม่รู้ปลายทาง สภาพแวดล้อมของฉาก แต่อย่างไรก็ตาม การเคลื่อนไหวก็พอที่จะจำลองได้โดยการใช้โมเดลความเร็วหรือรูปแบบความเร่งที่คงที่ในช่วงระยะเวลาสั้นๆ ดังนั้นการพิจารณาการเคลื่อนไหวในการเชื่อมโยงข้อมูล ควรพิจารณาในรูปแบบที่แตกต่างกันในช่วงระยะสั้นและระยะยาว และนี่เป็นแรงบันดาลใจให้งานวิจัย [2] นำแนวทางการทำงานเป็นลำดับขั้นมาประยุกต์ใช้ เช่น การเริ่มต้นด้วยการหาแตร็คเล็ตก่อน จากนั้นจึงรวมแตร็คเล็ตทั้งหมดให้เป็นเส้นทางการเดินที่สมบูรณ์

3.3.1 การหาแตร็คเล็ตโดยวิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด

แบ่งวิดีโอออกเป็นจำนวน s ส่วนย่อย (segment) แต่ละส่วนย่อยจำนวน f เฟรม งานวิจัย[2] ได้เสนอวิธีการเชื่อมโยงข้อมูลสำหรับการหาแตร็คเล็ตซึ่งเป็นการหาจากการเคลื่อนไหวและรูปพรรณสัณฐานที่สอดคล้องกันจากทั้งส่วนย่อยของวิดีโอ โดยข้อมูลนำเข้าสำหรับการเชื่อมโยงข้อมูลเพื่อหาแตร็คเล็ตอยู่ในรูปแบบกราฟ $G = (V, E, w)$ โดยกำหนดให้ V เป็นเซตของโหนด E เป็นเซตของขอบ และ w เป็นน้ำหนักของขอบ โดย V แบ่งออกเป็นจำนวน f กลุ่ม ซึ่งแต่ละกลุ่มหมายถึงหนึ่งเฟรม และโหนดที่อยู่ข้างในหมายถึงการตรวจจับบุคคลในเฟรมนั้นๆ ให้ C_i คือ เฟรมที่ i (\equiv กลุ่ม) โดย $i \in \mathbb{Z} : 1 \leq i \leq f$ และให้ v_m^i หมายถึง การตรวจจับบุคคลที่ m (\equiv โหนด) ในเฟรมที่ i ดังนั้น $C_i = \{v_1^i, v_2^i, v_3^i, \dots\}$ กำหนดให้ขอบของกราฟคือ $E = \{(v_m^i, v_n^j) | i \neq j\}$ ซึ่งมีความหมายว่าทุกโหนดใน G จะเชื่อมกันได้ครบเท่าที่แต่ละโหนดไม่ได้อยู่ในกลุ่มเดียวกัน v_m^i เป็นตัวแทนของคุณลักษณะตาแห่ง x_m^i เป็นพิกัด 2 มิติของตาแห่งตรงกลางของบุคคลที่ตรวจจับได้ และคุณลักษณะรูปพรรณสัณฐาน $\phi_{m,l}^i$ เป็นค่าฮิสโตแกรมสีของการแบ่งส่วนร่างตาแห่งที่ l ของการตรวจจับ v_m^i คำนวณน้ำหนักของขอบระหว่างสองโหนดจะเป็น $w : E \rightarrow \mathbb{R}^+$ ซึ่งคือค่าความคล้ายคลึงระหว่างการตรวจจับบุคคลสองบุคคล

$$w(v_m^i, v_n^j) = \sum_{l=1}^8 k(\phi_l^i, \phi_l^j) \quad (3.8)$$

โดย k หมายถึง การอินเตอร์เซกชันของฮิสโตแกรม (histogram intersection kernel)

งานในการหาแตร็คเล็ตของบุคคลคนเดียวกันในแต่ละส่วนย่อยของวิดีโอจำเป็นต้องมีการระบุตัวบุคคลที่ตรวจจับได้ในแต่ละเฟรม ดังนั้น คำตอบที่เป็นไปได้ของปัญหานี้จะถูกนำเสนอในรูปแบบของกราฟย่อยของกราฟ G ซึ่งเป็นกราฟที่มีหนึ่งโหนด (\equiv การตรวจจับ) ที่ถูกเลือกจากแต่ละกลุ่ม (\equiv เฟรม) กำหนดให้กราฟย่อยนี้เป็นคำตอบที่เป็นไปได้ในรูปแบบ $G_s = (V_s, E_s, w_s)$ ดังนั้น G_s จะเก็บเซตของโหนดที่นำเข้ามาจาก $V_s = \{v_a^1, v_b^2, v_c^3, \dots\}$ ซึ่งหมายถึง โหนดที่ a จากกลุ่มที่ 1 โหนดที่ b จากกลุ่มที่ 2 และโหนดอื่นที่ถูกเลือกจะอยู่ใน V_s จากคานิยามจะได้ $E_s = \{E(p, q) | p \in V_s, q \in V_s\}$ และ $w_s = \{w(p, q) | p \in V_s, q \in V_s\}$ จงจำไว้ว่าคำตอบที่เป็นไปได้ในรูปแบบ G_s ที่เป็นแตร็คเล็ต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

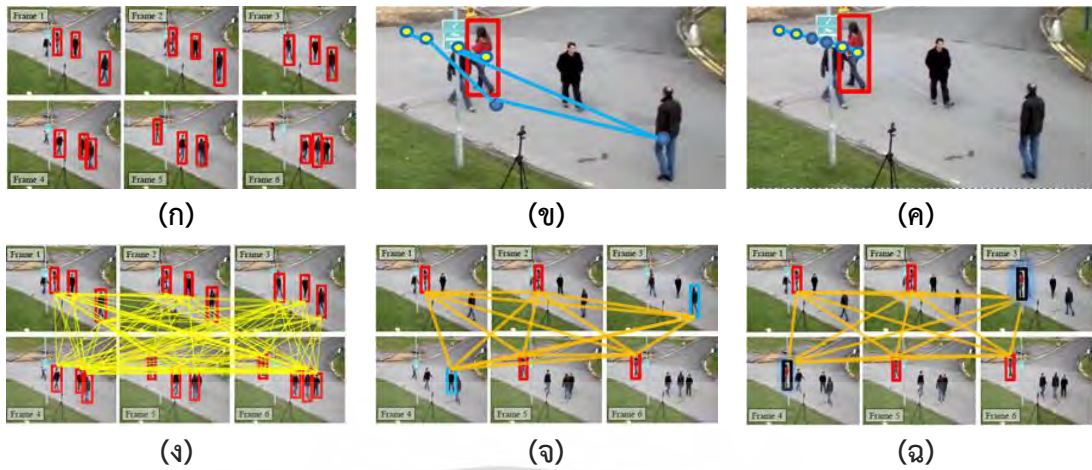
ของบุคคล 1 คน และไม่ใช่ทุกคนที่ปรากฏในส่วนย่อยของวิดีโอ นั้น จากรูปที่ 3.6 (ก) แสดงการตรวจจับบุคคลในส่วนย่อยของวิดีโอจำนวน 6 เฟรม พร้อมด้วยกราฟ G ที่แสดงใน (ง) โดยใน (จ) แสดงคำตอบที่เป็นไปได้ทั้งหมด และ G_s แสดงใน (ข) เนื่องจากเซตของโหนด V_s เพียงพอที่จะอธิบายรูปแบบของ G_s โดยหลังจากนี้เป็นต้นไป จะใช้ V_s เพื่อแสดงถึงคำตอบที่เป็นไปได้

กำหนดให้ค่ารูปพรรณสัณฐานของคำตอบที่เป็นไปได้ V_s คือ

$$Y_{appearance}(V_s) = \frac{1}{2} \left(\sum_{i=1}^f \sum_{j=1, j \neq i}^f w(V_s(i), V_s(j)) \right) \quad (3.9)$$

เป็นค่าของกราฟบริบูรณ์ (the complete graph) ซึ่งเกิดจากโหนดใน V_s จากสมการที่ 3.9 เป็นฟังก์ชันการคิดค่าโดยรวม เนื่องจากมีพื้นฐานอยู่บนการเปรียบเทียบเป็นคู่ๆ การตรวจจับบุคคลในคำตอบที่เป็นไปได้ โดยจะเทียบทุกคู่ที่มีโอกาส โดยไม่สนใจความใกล้ชิดกันของช่วงเวลา โดยอยู่บนพื้นฐานที่สมมติฐานว่าลักษณะรูปพรรณสัณฐานของคนจะไม่เปลี่ยนแปลงอย่างเห็นได้ชัดภายในส่วนย่อยของวิดีโอ การซ้อนกันของขอบเขตกรอบการตรวจจับ การถูกบดบัง การที่ภาพถูกรบกวน พื้นหลังที่รก และอื่นๆ จะส่งผลทำให้ในเส้นทางการเดินอาจเกิดระบุดับบุคคลสลับกันได้ทั้งในวิธีนี้ และโดยเฉพาะในวิธีจำกัดขอบเขตพื้นที่ในเชิงเวลา โดยจากสมการที่ 3.9 จะช่วยลดโอกาสของการระบุดับบุคคลสลับกันได้ เนื่องจากทุกคู่ของการตรวจจับที่เป็นไปได้จะเปรียบเทียบกันโดยไม่สนใจลำดับช่วงเวลา

จากการหาคำตอบที่เป็นไปได้โดยใช้ค่ารูปพรรณสัณฐานที่น้อยที่สุด นั่นคือ $\operatorname{argmin}_{V_s} (Y_{appearance}(V_s))$ ทำให้สามารถตรวจพบแตรัคเก็ตของบุคคลมีค่าฮิสโตแกรมสีที่มีความเสถียรในแต่ละส่วนย่อยของวิดีโอได้



รูปที่ 3.6 แสดงการหาแตร็คเคลื่อนจาก 6 เฟรมภายใน 1 ส่วนย่อยของวิดีโอ

(ก) การตรวจจับบุคคลใน ส่วนย่อยของวิดีโอจำนวน 6 เฟรม

(ข) คำตอบที่เป็นไปได้ในกรณีที่คนโดนบดบัง

(ค) คำตอบที่เป็นไปได้ในกรณีที่คนโดนบดบัง และมีการเพิ่มโหนดสมมุติ

(ง) กราฟ G (จ) กราฟคำตอบที่เป็นไปได้ในกรณีที่คนโดนบดบัง

(ฉ) กราฟคำตอบที่เป็นไปได้ในกรณีที่คนโดนบดบัง และมีการเพิ่มโหนดสมมุติ

ในปัญหาการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด โหนดของกราฟจะถูกจับกลุ่มในรูปที่เป็นกลุ่มแยกอิสระ วัตถุประสงค์เพื่อหาซับเซตของโหนดที่ประกอบด้วยเพียงหนึ่งโหนดจากแต่ละกลุ่ม โดยมีข้อจำกัดคือต้องมีค่าน้อยที่สุดตลอดช่วงกราฟของซับเซตโหนดนั้นด้วย

นิยามให้กราฟ $G = (V, E, w)$ โดย G เป็นกราฟที่คิดน้ำหนักและไม่มีการทิศทาง V เป็นเซตของโหนดทั้งหมด E เป็นเซตของขอบ และ $w : E \rightarrow \mathbb{R}^+$ เป็นน้ำหนักของขอบ โดยเซตของโหนด V จะถูกแบ่งออกเป็น f กลุ่ม C_1, C_2, \dots, C_f โดยที่ทุกกลุ่มจะต้องไม่มีสมาชิกร่วมกันเลยกล่าวคือ $C_1 \cup C_2 \cup \dots \cup C_f = V$ และ $C_i \cap C_j = \emptyset (1 \leq i \neq j \leq k)$ คำตอบที่เป็นไปได้ของ GMCP จะอยู่ในรูปกราฟย่อย $G_s = (V_s, E_s, w_s)$ โดย V_s เป็นซับเซตของ V ที่เป็นเพียงหนึ่งโหนดจากแต่ละกลุ่มเท่านั้น E_s เป็นซับเซตของ E ซึ่งเข้ามาจากโหนด V_s และ w_s เป็นน้ำหนักจาก w เป้าหมายของ GMCP คือการหาคำตอบที่เป็นไปได้โดยเลือกค่าน้อยที่สุด โดยค่าดังกล่าวคือค่าผลรวมของน้ำหนักทั้งหมดในกราฟย่อยของค คำตอบที่เป็นไปได้

ในการคำนวณนี้ จะต้องมีขอบอย่างน้อย 1 ขอบใน E ซึ่งมีไว้สำหรับเก็บคู่ที่เป็นไปได้ของโหนดใน V トラบเท่าที่โหนดเหล่านั้นไม่ได้อยู่ในกลุ่มเดียวกัน ดังนั้นกราฟย่อย G_s จึงมีความเป็นกราฟบริบูรณ์โดยสร้างมาจากค คำตอบที่เป็นไปได้ด้วยวิธี GMCP

ในการที่จะรวมคุณลักษณะการเคลื่อนไหวเข้ากับลักษณะรูปพรรณสัณฐานเพื่อใช้ในการเชื่อมโยงข้อมูล จะต้องเพิ่มพจน์อีก 1 พจน์ เข้าไปในฟังก์ชันและกำหนดให้ฟังก์ชันนี้ในการเชื่อมโยงข้อมูลทั้งหมด โดยฟังก์ชันมีดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\hat{V}_s = \operatorname{argmin}_{V_s} (\gamma_{\text{appearance}}(V_s) + \alpha \cdot \gamma_{\text{motion}}(V_s)) \quad (3.10)$$

โดย \hat{V}_s เป็นคำตอบที่เหมาะสมที่สุดของการเชื่อมโยงสำหรับหนึ่งแทร็คเล็ต และ α คือค่าคงที่ของการผสมเพื่อทำให้การรวมคุณลักษณะรูปพรรณสัณฐานและลักษณะการเคลื่อนไหวมีความสมดุลกัน

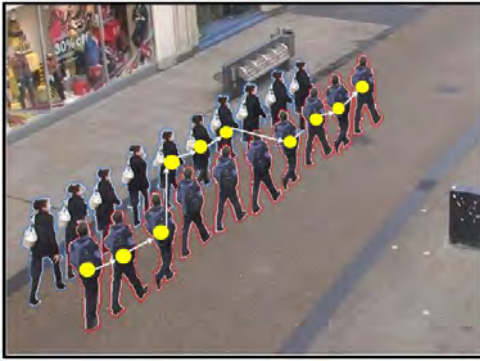
การทำ \hat{V}_s จากในสมการที่ 3.10 จะให้ผลลัพธ์เป็นแทร็คเล็ตของคนหนึ่งคนในส่วนของวิดีโอ ดังนั้นในการที่จะหา แทร็คเล็ตของคนในส่วนของวิดีโอ จะต้องคำนวณตามสมการที่ 3.10 ซ้ำอีกหลายครั้ง โดยในครั้งแรกที่สมการที่ 3.10 คำนวณเสร็จ อัลกอริทึมจะหาแทร็คเล็ตที่มีค่ารวมน้อยที่สุด ซึ่งนั่นคือ รูปแบบคุณลักษณะรูปพรรณสัณฐานและลักษณะการเคลื่อนไหวมีความเสถียรที่สุด จากนั้น แทร็คเล็ตที่ดีที่สุดที่ถูกเลือกแล้วใน \hat{V}_s จะถูกคัดออกมาจาก G และทำการหาแทร็คเล็ตซ้ำสำหรับของคนถัดไป และจะทำซ้ำจนกว่าใน G จะเหลือจำนวน 0 โหนดหรือเหลือแต่โหนดที่ใช้ไม่ได้

รูปแบบค่าคุณลักษณะการเคลื่อนไหวเพื่อการทำแทร็คเล็ต (Tracklet - global motion cost model) ในลำดับของการนำค่าการเคลื่อนไหว เพื่อไปใช้ในการคำนวณในสมการที่ 3.10 ต้องคำนวณค่าโดยอยู่บนพื้นฐานของการเคลื่อนไหวเพื่อใช้หาคำตอบที่เป็นไปได้ V_s โดยเวกเตอร์ความเร็วเชิงตำแหน่งของคำตอบที่เป็นไปได้ V_s กำหนดให้อยู่ในรูป $\dot{X}_s(i) = X(i+1) - X(i)$ โดย $1 \leq i \leq (f-1)$ วิธีการหนึ่งที่สามารถใช้คำนวณค่าการเคลื่อนไหวนั้นคือการคำนวณค่าความเบี่ยงเบนจากรูปแบบที่สันนิษฐาน เช่น ใช้ความเร็วคงที่ ซึ่งสามารถทำได้โดยใช้เวกเตอร์ความเร็วของแต่ละคนมาทำการทำนายตำแหน่งของการตรวจจับถัดไป หลังจากนั้นก็ทำการหาค่าความผิดพลาดระหว่างตำแหน่งที่ได้จากการทำนายกับตำแหน่งในชุดคำตอบที่สามารถเป็นไปได้ ซึ่งวิธีการแบบนี้ได้ใช้เป็นหลักในวิธีการจับคู่สองฝ่ายของกราฟและในวิธีการอื่นๆอีก [5,6] แต่อย่างไรก็ตาม ในขอบเขตงานวิจัย [2] หนึ่งคำตอบที่เป็นไปได้นั้นหมายถึงหนึ่งแทร็คเล็ตตลอดทั้งส่วนของวิดีโอ ดังนั้นจึงสามารถคำนวณค่าการเคลื่อนไหวได้อย่างมีประสิทธิภาพ ซึ่งสามารถมั่นใจได้ว่ารูปแบบนี้ทำให้ผลลัพธ์มีความต่อเนื่องและสอดคล้องกัน ในการคำนวณนั้นได้ตั้งข้อสันนิษฐานว่ารูปแบบการเดินนั้นมีความเร็วที่คงที่ตลอดในส่วนย่อยของวิดีโอและจะสามารถคำนวณคุณลักษณะการเคลื่อนไหวได้จาก :

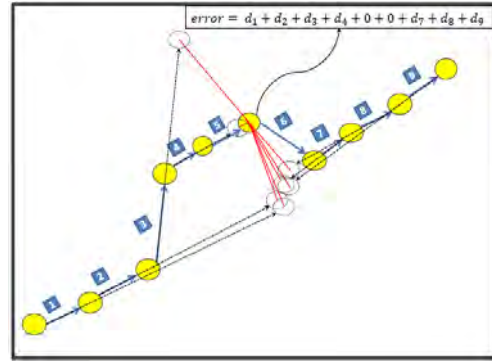
$$\gamma_{\text{motion}}(V_s) = \sum_{i=1}^s \sum_{j=1}^{s-1} \frac{\overbrace{|X_s(i) - [X_s(j) + \dot{X}_s(j) \cdot (i-j)]|}^{\text{deviation}}}{\text{prediction}} \quad (3.11)$$

จากสมการที่ 3.11 พจน์ในวงเล็บคือการทำนายตำแหน่งของโหนด $V_s(i)$ โดยใช้ $\dot{X}_s(j)$ ในสมการที่ 3.11 ได้ตั้งสันนิษฐานว่าบุคคลเคลื่อนที่ด้วยความเร็วคงที่ในหนึ่งส่วนของวิดีโอ และในแต่ละเวกเตอร์ \dot{X}_s จะถูกใช้ในการทำนายตำแหน่งของโหนดอื่นๆ ทั้งหมดในคำตอบที่เป็นไปได้ V_s

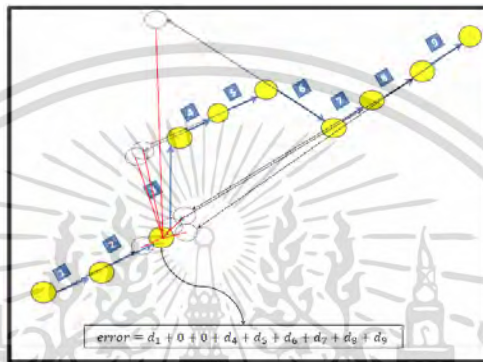
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)



(ค)

รูปที่ 3.7 คุณลักษณะการเคลื่อนไหวเพื่อการหาแตร็คเล็ต

(ก) คำตอบที่เป็นไปได้ในกรณีที่เกิดการติดตามสลับบุคคล

(ข) การคิดค่าการเคลื่อนไหวที่โหนดที่ 6

(ข) การคิดค่าการเคลื่อนไหวที่โหนดที่ 3

จากรูปที่ 3.7 แสดงรายละเอียดดังต่อไปนี้ ในรูป (ก) แสดงคำตอบที่เป็นไปได้ซึ่งเป็นของบุคคลในภาพที่มีขอบสีแดง แต่อย่างไรก็ตามมี 3 การตรวจจับของบุคคลอื่นซึ่งถือว่าเป็นการเลือกที่ผิดพลาดในคำตอบที่เป็นไปได้นี้ ซึ่งคาดการณ์ว่า 3 การตรวจจับที่เลือกมาผิดนั้นจะเพิ่มค่าการเคลื่อนไหวขึ้นมาก ขณะที่โหนดที่เลือกตัวอื่นๆ ที่สอดคล้องจะเพิ่มค่าการเคลื่อนไหวเล็กน้อย ค่าจากสมการที่ 3.11 แสดงสำหรับสองโหนดคือ $i = 6$ และ $i = 3$ ในส่วนรูป (ข) และ (ค) ตามลำดับวงกลมสีแดงแสดงถึงตำแหน่งที่ทำการทนายโหนด i ส่วนเส้นสีแดงอธิบายถึงระยะห่างระหว่างตำแหน่งที่ทนายกับตำแหน่งของ $X_s(i)$ ซึ่งเป็นการแสดงถึงค่าความเบี่ยงเบนจากรูปแบบ ค่าของโหนด i นั้นจะเพิ่มค่าการเคลื่อนไหวเพื่อไปรวมเป็นผลรวมความต่าง ซึ่งจากที่เห็น โหนด $i = 6$ ที่ไม่ได้อยู่ในวิถีแตร็คเล็ตหลัก ทำให้มีค่าการเคลื่อนไหวที่มาก ขณะที่โหนด $i = 3$ จะเพิ่มค่าเพียงแค่น้อย

ดังนั้น ในทางตรงกันข้าม รูปแบบการเคลื่อนไหวที่ต่อเนื่อง ค่าการเคลื่อนไหวในสมการที่ 3.11 จะคำนวณจากการวัดค่าเบี่ยงเบนจากรูปแบบความเร็วคงที่ในภาพรวมของแตร็คเล็ต เพราะทุกๆ โหนดมีผลต่อการคิดค่าการเคลื่อนไหวในโหนดอื่นๆ นอกเหนือจากความเร็วคงที่แล้วตามสมการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่ 3.11 ยังสมมุติฐานว่าความเร่งก็คงที่ด้วยและอนุพันธ์การเปลี่ยนแปลงระดับสูงก็คงที่เช่นกันเพื่อให้การพิจารณาเหตุการณ์ที่มีความซับซ้อนได้อย่างตรงไปตรงมา

จัดการปัญหาการโหนดบังด้วยโหนดสมมุติ (Hypothetical Nodes) ในบางกรณีเฟรมที่พิจารณาอาจไม่มีผลการตรวจจับบุคคลเนื่องจากคนๆนั้นโหนดบัง การตรวจจับผิดพลาด และสาเหตุอื่นๆ ซึ่งในประเด็นนี้สามารถรับมือได้ด้วยการเพิ่มโหนดสมมุติไปสู่แต่ละกลุ่ม ดังนั้นถ้าเฟรมใดเฟรมหนึ่งไม่มีการตรวจจับที่เหมาะสม โหนดสมมุติจะเป็นผู้ถูกเลือกแทน

จำเป็นต้องมีการกำหนดค่าคุณลักษณะรูปพรรณสัณฐานและค่าคุณลักษณะการเคลื่อนไหวทั้งสองค่า ให้กับโหนดสมมุติ ทุกโหนดใน G จากนั้นการแก้ปัญหาเพื่อหาค่าที่เหมาะสมจะเป็นกระบวนการที่ 3.10 ในแต่ละรอบทั้งสองค่านี้จะถูกคำนวณใหม่ให้กับโหนดสมมุติ ซึ่งในส่วนของการคำนวณและการปรับปรุงโหนดสมมุติจะขออธิบายในส่วนถัดไป

V_s ถูกคาดหวังให้รวบรวมการตรวจจับของบุคคลหนึ่งคน (ค่าที่อยู่ในเกณฑ์ที่ถูกต้อง (inlier)) และการตรวจจับอื่นๆ (ค่าที่ผิดปกติ (outlier)) ในเฟรมต่างๆ ที่ไม่สามารถตรวจจับของคนๆนั้นได้ เนื่องจากส่วนย่อยของวิดีโอเป็นช่วงเวลาสั้นๆ จึงสามารถตั้งสมมุติฐานว่าการเคลื่อนไหวของคนจะเคลื่อนไหวด้วยความเร็วคงที่ตลอดทั้งส่วนย่อยของวิดีโอและจะใช้สมมุติฐานนี้ในการแยกแยะว่าเป็นค่าที่อยู่ในเกณฑ์หรือค่าที่ผิดปกติใน V_s อีกด้วย ตามที่การเคลื่อนไหวด้วยความเร็วคงที่นั้นอยู่ในรูปฟังก์ชันสองมิติ ตาแหน่งของการตรวจจับสามารถแสดงขึ้นโดยใช้ $X_s(i) = a_1 i + a_0$ โดยที่ a_1 และ a_0 เป็นเวกเตอร์คงที่ในระบบสองมิติ ดังนั้นสามารถระบุค่าที่อยู่ในเกณฑ์และค่าที่ผิดปกติด้วยกำหนดรูปแบบความเร็วคงที่จาก a_1 และ a_0 โดยใช้

$$V_s^{inliers} = \{V_s(i) : |a_1 i + a_0 - X_s(i)| < \delta\} \quad (3.12)$$

โดยที่ δ คือค่าเงื่อนไขการเบี่ยงเบนที่เหมาะสม โดยตัวแปรที่ดีที่สุดของโมเดลความเร็วคงที่ของแทร็กเล็ตต้องเป็นรูปแบบซึ่งจะหาให้ได้ จำนวนแทร็กเล็ตที่อยู่ในเกณฑ์มากที่สุด

$$\hat{a}_1, \hat{a}_0 = \underset{a_1, a_0}{\operatorname{argmax}} (\# \{V_s^{inliers}\}) \quad (3.13)$$

ให้ $\#$ หมายถึง จำนวนสมาชิกของเซต เนื่องจาก V_s ประกอบด้วยค่าที่อยู่ในเกณฑ์และค่าที่ผิดปกติ งานวิจัย[2] ใช้เทคนิค RANSAC เพื่อจะได้ค่า a_1 และ a_0 ตามในสมการที่ 3.13 และไปหาเกณฑ์ความผิดพลาดต่อในสมการที่ 3.12

โหนดที่อยู่ในเกณฑ์ใน V_s คือโหนดที่ผ่านการคัดเลือกจากสมการที่ 3.12 โดยใช้ \hat{a}_1 และ \hat{a}_0 โดยตำแหน่งพิกัดของโหนดสมมุติเป็นการคำนวณจากสมการ $x_H^i = \hat{a}_1 i + \hat{a}_0$ โดย x_H^i หมายถึงพิกัดของโหนดสมมุติของกลุ่มที่ i สำหรับค่าคุณลักษณะรูปพรรณสัณฐานของโหนดสมมุติจะเป็น

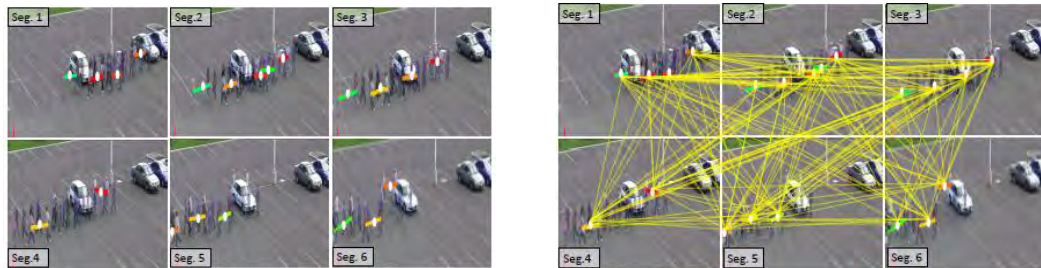
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าเฉลี่ยของคุณลักษณะรูปพรรณสัณฐานของโหนดที่อยู่ในเกณฑ์ (เป็นค่าเฉลี่ยของฮิสโตแกรมสี) แต่อย่างไรก็ตาม ค่าปรับคังที่จะถูกเพิ่มกับค่าน้ำหนักของเส้นขอบที่เชื่อมไปยังโหนดสมมุติใน G แต่จะพยายามหลีกเลี่ยงการเลือกโหนดสมมุติถ้าหากในเฟรมนั้นมีการตรวจจับที่เหมาะสมอยู่

จากที่ได้กล่าวมาก่อนนี้ โหนดสมมุติจะถูกปรับคังเมื่อตอนจบการคำนวณสมการที่ 3.10 ในแต่ละรอบ ในช่วงรอบแรกๆนั้น โหนดสมมุติจะยังไม่ถูกเลือกจากอัลกอริทึม อัลกอริทึมจะทำการเลือกการตรวจจับอื่นๆก่อน เมื่อการคำนวณผ่านไปสักระยะหนึ่ง ในกลุ่มซึ่งมีการตรวจจับที่ถูกต้องอยู่ จะเริ่มถูกเลือกจนหมด และโหนดสมมุติจะเริ่มมีหน้าที่เข้ามาสนับสนุน จนกระทั่งอัลกอริทึมเข้าสู่คาตอบสุดท้าย V_r รูปที่ 3.6 (ฉ) แสดง V_r มีการเลือกใช้โหนดสมมุติสองโหนดเพื่อทดแทนเฟรมที่บุคคลโดนบดบัง ใน (ค) แสดงถึงรูปแบบเส้นทางการเดิน V_r

3.3.2 การรวมแตร็คเล็ดเป็นเส้นทางการเดินโดยใช้การพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด

จากที่ได้อธิบายไปก่อนหน้านี้ สามารถแบ่งวิดีโอให้กลายเป็น s ส่วนย่อย และหาแตร็คเล็ดของคนเดินทุกๆคนในแต่ละส่วนย่อยของวิดีโอโดยใช้วิธีการที่ได้อธิบายไปแล้วในส่วนที่ 3.3.1 ในส่วนของการสร้างเส้นทางการเดินของบุคคลที่สมบูรณ์จนครบทั้งวิดีโอ นั้น ต้องการการรวมแตร็คเล็ดจากของแต่ละบุคคล ซึ่งการเชื่อมโยงข้อมูลแบบนี้สามารถใช้ได้หลากหลายวิธี ยกตัวอย่างเช่น การจับคู่สองฝ่ายของกราฟ [5,6] แต่อย่างไรก็ตาม ในขอบเขตการทางานของงานวิจัย[2] จะใช้ GMCP เหมือนเดิมส สำหรับขั้นตอนการหาแตร็คเล็ดและรวมเข้าด้วยกัน ดังนั้น กำหนดให้กลุ่มและโหนดใน G' จะหมายถึงส่วนย่อยของวิดีโอและแตร็คเล็ดตามลำดับ (แทนที่กลุ่มและโหนดหมายถึงเฟรมและการตรวจจับบุคคลตามลำดับในส่วนที่ 3.3.1) คุณลักษณะรูปพรรณสัณฐานของโหนดที่เป็นตัวแทนของแตร็คเล็ดจะเป็นค่าที่ได้จากการเฉลี่ยค่ารูปพรรณสัณฐานของการตรวจจับบุคคลในแตร็คเล็ดนั้นๆ (เป็นค่าเฉลี่ยของฮิสโตแกรมสี) และ x'_m กำหนดให้เป็นตำแหน่งตรงกลางของแตร็คเล็ด จากรูปที่ 3.8 (ก) แสดงถึงแตร็คเล็ดจาก 6 ส่วนย่อยของวิดีโอพร้อมรูปภาพบริบูรณ์

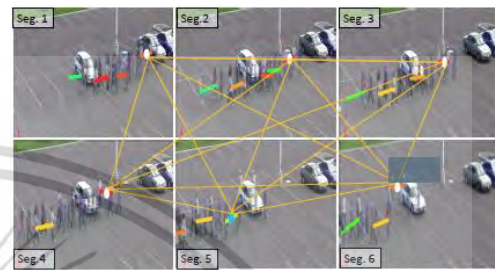


(ก)

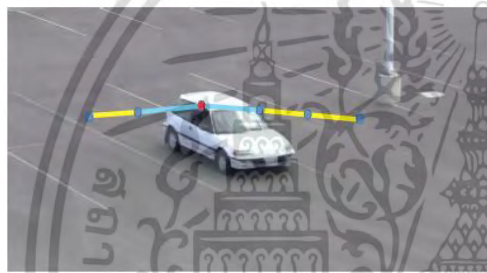
(ข)



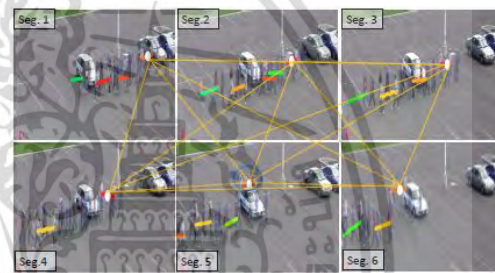
(ค)



(ง)



(จ)



(ฉ)

รูปที่ 3.8 การรวมแทร็คเล็กลงเป็นเส้นทางการเดินในกรณีต่างๆ

(ก) แทร็คเล็กลงจากแต่ละส่วนย่อยของวิดีโอ

(ข) กราฟ G' (ค) คำตอบที่เป็นไปได้โดยที่ยังไม่เพิ่มโหนดสมมุติ

(ง) การเลือกคำตอบที่เป็นไปได้โดยที่ยังไม่เพิ่มโหนดสมมุติ

(จ) เส้นทางการเดินที่สมบูรณ์

(ฉ) การเลือกคำตอบที่เป็นไปได้ที่มีการเพิ่มโหนดสมมุติ

มีข้อควรทราบว่าการเชื่อมโยงข้อมูลในระดับการหาเส้นทางการเดินนั้นมีความแตกต่างจากระดับในการหาแทร็คเล็กลง จากที่ได้ตั้งข้อสมมุติฐานว่าการเดินของคนจะมีความเร็วคงที่ในหนึ่งส่วนย่อยของวิดีโอ แต่ในการจำลองการเคลื่อนไหวของคนในระยะเวลานานนั้น ในการติดตามจะมีความยากและซับซ้อนมากยิ่งขึ้น ซึ่งโดยทั่วไป มันเป็นเรื่องยากที่จะจำลองการเคลื่อนไหวของคนเดินในช่วงเวลานานโดยปราศจากความรู้ที่เกี่ยวกับโครงสร้างของฉาก ตำแหน่งปลายทาง รูปแบบปฏิสัมพันธ์ของผู้คน และอื่นๆ ซึ่งด้วยปัญหาที่แตกต่างเหล่านี้ จึงไม่สามารถใช้สมการการคำนวณค่าคุณลักษณะการเคลื่อนไหวที่ได้อธิบายไว้ใน 3.3.1 และไม่สามารถใช้วิธีการคำนวณโหนดสมมุติที่ได้อธิบายไว้ใน 3.3.1 เช่นกัน อย่างไรก็ตาม ก็ยังมุ่งหวังว่าเส้นทางการเดินที่สมบูรณ์นั้นจะต้องมีความเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ราบรื่นในเชิงเวลา ดังนั้นในระดับการคำนวณนี้ มีความจำเป็นที่จะต้องเปลี่ยนวิธีการคำนวณค่าคุณลักษณะการเคลื่อนไหวและการคำนวณโหนดสมมุติเป็นการพิจารณาความต่อเนื่องในแต่ละจุดข้อมูล แทนการมองภาพโดยรวมทั้งหมด ซึ่งจะขอกล่าวดังต่อไปนี้

ค่าการเคลื่อนไหว (Motion-Cost) ในการคำนวณค่าการเคลื่อนไหวสำหรับคำตอบที่เป็นไปได้ V'_s จะใช้เวกเตอร์ความเร็วของแต่ละโหนดก่อนหน้ามาใช้ในการคำนวณ โดยการขยายออกมาอย่างต่อเนื่องเป็นช่วง

$$Y'_{motion}(V'_s) = \sum_{i=3}^s |X'_s(i) - [X'_s(i-2) + 2X'_s(i-2)]| \quad (3.14)$$

เปรียบเทียบกับสมการที่ 3.14 กับสมการที่ 3.11 ที่ใช้แนวทางวิเคราะห์โดยภาพรวม จะเห็นว่าในสมการที่ 3.14 ได้ตัดทิ้งผลรวมตัวที่สองในสมการที่ 3.11 เพื่อให้สามารถนำผลของเส้นทางการเดินก่อนหน้ามาร่วมพิจารณาในการคำนวณค่าการเคลื่อนไหวได้

รับมือปัญหาแตร็คเล็ทที่โดนบดบังโดยใช้โหนดสมมุติ ในระดับของแตร็คเล็ท เหตุการณ์ที่โดนบดบังในระยะสั้นและการตรวจจับที่ผิดพลาดได้มีการรับมือกับปัญหาแล้วด้วยการใช้โหนดสมมุติ แต่อย่างไรก็ตาม ถ้าหากเกิดเหตุการณ์ที่คนเดินถูกบดบังเป็นระยะเวลาที่นานกว่าส่วนย่อยของวิดีโอเช่น เกินกว่า 50 เฟรม ซึ่งนั่นจะทำให้บุคคลคนนั้นจะไม่มีแตร็คเล็ทในส่วนย่อยของวิดีโออื่นๆ ซึ่งเหตุการณ์นี้นั้นจะนำไปสู่เหตุการณ์แตร็คเล็ทถูกบดบัง ในการรับมือกับปัญหานี้ จะต้องเพิ่มโหนดสมมุติเข้าไปในกลุ่มของแตร็คเล็ทของกราฟความสัมพันธ์ G' ด้วยการทานายตำแหน่งด้วยวิธีการคาดการณ์ตำแหน่ง

$$x'_{H_f} = |X'_s(i-2) + 2X'_s(i-2)| \quad (3.15)$$

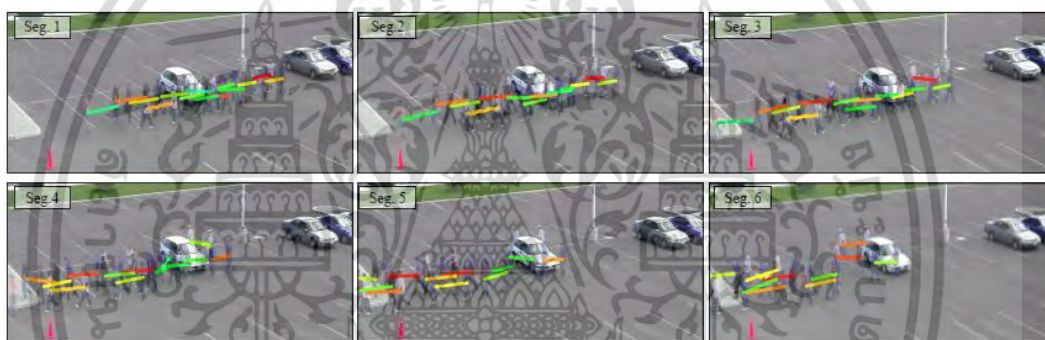
จากสมการที่ 3.15 จะเห็นว่าใช้เพียงเส้นทางการเดินก่อนหน้ามาใช้ในการคำนวณตำแหน่งของโหนดสมมุติ แทนที่จะใช้วิธีการพิจารณาภาพรวมเช่นที่ใช้ในส่วนที่ 3.3.1 ซึ่งเป็นการเอาทุกโหนดมาใช้

ในระดับของการเชื่อมโยงแตร็คเล็ท สามารถเพิ่มโหนดสมมุติจนวนสองโหนดไปยังแต่ละกลุ่มแทนที่จะใช้แค่โหนดเดียว กำหนดให้ H_f คือโหนดสมมุติข้างหน้า ซึ่งเกิดจากการนำเส้นทางการเดินขึ้นก่อนหน้ามาทานาย และ ให้ H_b คือโหนดสมมุติข้างหลัง ซึ่งเป็นการนำเส้นทางการเดินข้างหลังมาทานายขึ้นก่อนหน้า โดยโหนดสมมุติข้างหลังเป็นสิ่งจำเป็นในกรณีที่ตอนเริ่มต้นขึ้นมาแตร็คเล็ทในเส้นทางการเดินถูกบดบัง ซึ่งกรณีหนึ่งก็คือไม่มีแตร็คเล็ทตรงที่ก่อนจะเริ่มเส้นทางการเดิน จึงจำเป็นต้องมีการทานายตำแหน่งของโหนดสมมุติตรงจุดที่ไม่มีแตร็คเล็ทนั้น โหนดสมมุติข้างหลังจึงถูกเพิ่มเข้าไปซึ่งได้มาจากการนำส่วนปลายของเส้นทางการเดินมาทานายทานายตำแหน่งของโหนดสมมุติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตรงจุดเริ่มต้นที่ไม่มีแทร็คเล็ท ดังนั้น ตำแหน่งของโหนดสมมุติข้างหลังจะสามารถคำนวณได้จากสมการ $x'_{H_b} = |X'_s(i+2) + 2X'_s(i+1)|$

คุณลักษณะรูปพรรณสัณฐานของโหนดสมมุติก่อนหน้าและโหนดสมมุติข้างหลังจะเป็นเหมือนกับโหนดที่ใช้ทำนายตำแหน่งของโหนดสมมุติ นั่นคือ $v'_s(i-2)$ สำหรับโหนดสมมุติก่อนหน้า และ $v'_s(i+2)$ สำหรับโหนดสมมุติข้างหลัง โหนดสมมุติที่ถูกเพิ่มเข้าในระดับของการรวมแทร็คเล็ทนั้นสามารถนำมาใช้ประโยชน์เพื่อแก้ปัญหาคนเข้าหรือออกระหว่างในวิดีโอ ยกตัวอย่างเช่น ถ้ามีคนเดินออกไปจากขอบเขตการมองเห็นก่อนที่จะจบวิดีโอ หรือมีคนเดินเข้ามาหลังจากที่ได้เริ่มวิดีโอไปแล้ว นั้นย่อมหมายถึงจะมีบางส่วนย่อยของวิดีโอที่ไม่มีแทร็คเล็ทของบุคคลเหล่านั้น ซึ่งสำหรับกรแก้ปัญหาในกรณีนั้นจะต้องมีการเลือกโหนดสมมุติให้กับส่วนย่อยของวิดีโออื่นๆ แต่อย่างไรก็ตามผลลัพธ์ตำแหน่งของโหนดสมมุติที่ได้จะเป็นบริเวณนอกขอบเขตการมองเห็นของเฟรม ซึ่งนั่นจะสอดคล้องกับตำแหน่งก่อนที่จะเดินเข้ามาในวิดีโอ หรือตำแหน่งหลังจากที่เดินออกไปจากวิดีโอแล้วนั่นเอง



(ก)



(ข)

รูปที่ 3.9 การเชื่อมโยงแทร็คเล็ทเป็นเส้นทางการเดิน

(ก) แทร็คเล็ทจาก 6 ส่วนย่อยของวิดีโอ

(ข) ผลลัพธ์เส้นทางการเดินจากการเชื่อมโยงแทร็คเล็ทของทุกส่วนย่อยของวิดีโอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การถดถอยเชิงเส้น (Linear Regression)

การวิเคราะห์การถดถอย (Regression analysis) สามารถใช้ในการหารูปแบบความสัมพันธ์ระหว่างหนึ่งหรือหลายตัวแปรทำนาย (Predictor variable) กับตัวแปรตอบสนอง (Response variable) โดยทั่วไปค่าของตัวแปรทำนายจะหาค่า ส่วนตัวแปรตอบสนองเป็นตัวแปรที่ต้องการการคาดการณ์โดยใช้ตัวแปรทำนายมาอธิบาย การวิเคราะห์การถดถอยเป็นทางเลือกที่ดีเมื่อตัวแปรทำนายมีลักษณะที่เป็นค่าต่อเนื่อง หลากหลายปัญหาสามารถแก้ได้ด้วยการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) หรือเรียกว่า การวิเคราะห์การถดถอยเส้นตรง (Straight-line regression analysis)

การวิเคราะห์การถดถอยเชิงเส้น จะประกอบด้วย ตัวแปรตอบสนอง y และตัวแปรทำนาย x โดยสามารถจัดให้อยู่ในรูปสมการถดถอยด้วยการจำลอง y เป็นสมการเชิงเส้นของเส้นของ x ดังสมการที่ 3.16

$$y = w_0 + w_1x \quad (3.16)$$

จากสมการ 3.16 กำหนดให้ w_0 และ w_1 เป็นค่าสัมประสิทธิ์การถดถอยที่กำหนดจุดตัดแกน Y และค่าความชันของเส้นตรง ค่าสัมประสิทธิ์การถดถอยสามารถคำนวณได้จากวิธีการกำลังสองน้อยที่สุด (method of least squares) ซึ่งเป็นวิธีการประมาณเส้นตรงที่เหมาะสมที่สุดที่จะช่วยลดความคลาดเคลื่อนระหว่างผลข้อมูลจริงกับค่าที่ได้จากการประมาณเส้นตรง กำหนดให้ D คือชุดฝึกสอนที่ประกอบไปด้วยตัวแปรทำนาย x และตัวแปรตอบสนอง y กลุ่มชุดฝึกสอนจะมีทั้งหมด $|D|$ ชุด สามารถเขียนให้อยู่ในรูป $(x_1, y_1), (x_2, y_2), \dots, (x_{|D|}, y_{|D|})$ ค่าสัมประสิทธิ์การถดถอยสามารถประมาณค่าได้จากสมการที่ 3.17 และ 3.18

$$w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2} \quad (3.17)$$

$$w_0 = \bar{y} - w_1\bar{x} \quad (3.18)$$

จากสมการที่ 3.17 และ 3.18 กำหนดให้ \bar{x} คือค่าเฉลี่ยของ $x_1, x_2, \dots, x_{|D|}$ และ \bar{y} คือค่าเฉลี่ยของ $y_1, y_2, \dots, y_{|D|}$

3.5 การปรับเรียบข้อมูลแบบถดถอยโดยพิจารณาเฉพาะส่วน

(Local Regression Smoothing)

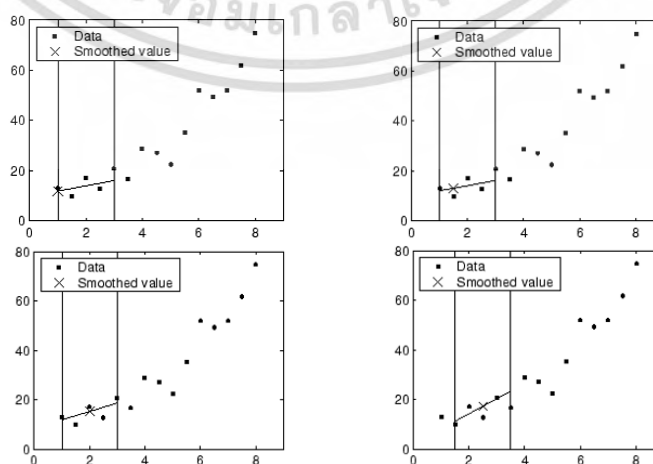
การปรับเรียบข้อมูลโดยพิจารณาเฉพาะส่วน แต่ละผลการปรับเรียบข้อมูลจะพิจารณาจากจุดข้อมูลบริเวณข้างเคียงภายในระยะช่วงข้อมูลที่กำหนด ในขั้นตอนการปรับเรียบมีการให้ค่าน้ำหนัก เนื่องจากค่าน้ำหนักของสมการถดถอยจะถูกคำนวณเพื่อจุดข้อมูลที่อยู่ในช่วง นอกจากฟังก์ชันถ่วงน้ำหนักของสมการถดถอย ยังมีสมการถดถอยโดยพิจารณาเฉพาะส่วนแบบแกร็ง ซึ่งวิธีการนี้จะมี ความทนทานต่อ outlier

3.5.1 สมการถดถอยโดยพิจารณาเฉพาะส่วน (Local Regression)

การปรับเรียบข้อมูลด้วยสมการถดถอยโดยพิจารณาเฉพาะส่วนในแต่ละจุดข้อมูล ในขั้นตอนแรก จากสมการที่ 3.19 จะต้อง คำนวณค่าน้ำหนักถดถอยสำหรับแต่ละจุดข้อมูลที่อยู่ในช่วง (span) ที่พิจารณา โดยค่าน้ำหนักสามารถคำนวณได้จากฟังก์ชัน tricube

$$w_i = \left(1 - \left|\frac{x - x_i}{d(x)}\right|^3\right)^3 \quad (3.19)$$

จากสมการที่ 3.19 กำหนดให้ x คือตัวแปรทำนายที่สัมพันธ์กับตัวแปรตอบสนองที่จะปรับเรียบ x_i คือค่าตำแหน่งใกล้เคียงกับค่า x ที่อยู่ในช่วงข้อมูลเดียวกัน และ $d(x)$ คือจำนวนพิกัดที่ห่างที่สุดจาก x ไปยังจุดที่ไกลที่สุดของตัวแปรทำนายที่อยู่ภายในช่วงเดียวกัน สำหรับค่าน้ำหนักถดถอยที่คำนวณได้ ถ้าไปใช้ในวิธี lowess สมการถดถอยจะใช้ฟังก์ชันพหุนามดีกรีที่ 1 ในการปรับเรียบข้อมูล และถ้าไปใช้ในวิธี loess สมการถดถอยจะใช้ฟังก์ชันพหุนามดีกรีที่ 2 ในการปรับเรียบข้อมูล โดยค่าปรับเรียบข้อมูลแต่ละตำแหน่งสามารถคำนวณได้จากค่าน้ำหนักถดถอยในแต่ละตำแหน่ง



รูปที่ 3.10 การปรับเรียบข้อมูลด้วยวิธี lowess

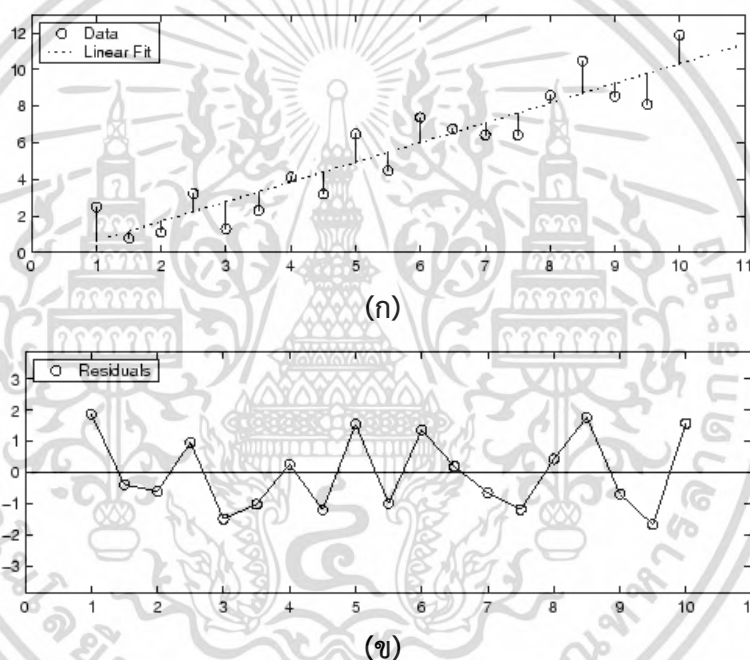
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.2 สมการถดถอยโดยพิจารณาเฉพาะส่วนแบบแกร่ง (Robust Local Regression)

สมการถดถอยโดยพิจารณาเฉพาะส่วนแบบแกร่ง เริ่มต้นขึ้นตอนจากการคำนวณค่าส่วนเหลือ (Residual Analysis)

$$r = y - \hat{y} \quad (3.20)$$

จากสมการที่ 3.20 กำหนดให้ y คือตัวแปรตอบสนอง และ \hat{y} คือตัวแปรตอบสนองที่ได้จากการทำนาย ดังนั้นค่าส่วนเหลือที่คำนวณจากโมเดลที่เหมาะสมคือค่าระยะห่างระหว่างตัวแปรตอบสนองและค่าที่เหมาะสมของตัวแปรตอบสนองที่เกิดจากการทำนายค่า



รูปที่ 3.11 การคำนวณค่าส่วนเหลือ

(ก) การคำนวณระยะห่างระหว่างชุดข้อมูลกับผลการทำนายรูปแบบเส้นตรง

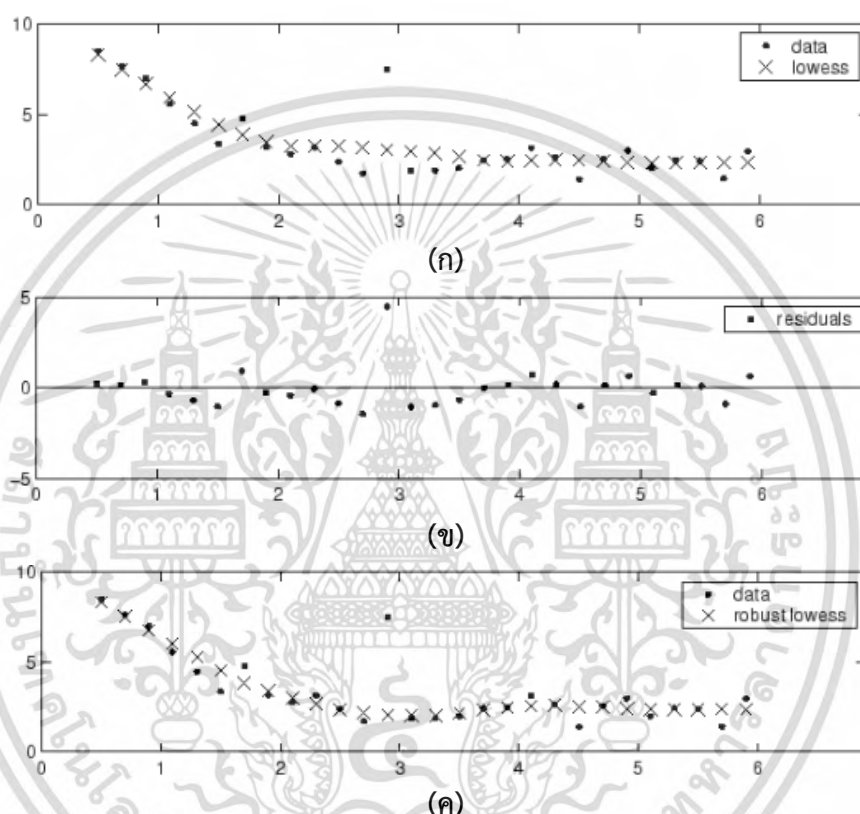
(ข) กราฟความสัมพันธ์ค่าส่วนเหลือบนรูปแบบเส้นศูนย์ (zero line)

ขั้นตอนถัดมาคือการคำนวณค่าถ่วงน้ำหนักแบบแกร่งสำหรับแต่ละจุดข้อมูลภายในช่วงโดยค่าน้ำหนักคิดด้วยฟังก์ชันไบสแควร์ (bisquare function)

$$w_i = \begin{cases} (1 - (r_i/6MAD)^2)^2, & |r_i| < 6MAD, \\ 0, & |r_i| \geq 6MAD, \end{cases} \quad (3.21)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้ r_i คือค่าส่วนเหลือของจุดข้อมูลที่ i ที่คำนวณจากการปรับเรียบข้อมูล และ MAD คือค่ามัธยฐานความผิดพลาดสัมบูรณ์ (median absolute deviation) ของค่าส่วนเหลือ โดยค่ามัธยฐานความผิดพลาดสัมบูรณ์จะเป็นการวัดการแพร่กระจายของค่าส่วนเหลือ โดยถ้า r_i มีค่าน้อยกว่า $6MAD$ ค่าน้ำหนักแบบถ่วงจะมีค่าเข้าใกล้ 1 ถ้า r_i มีค่ามากกว่า $6MAD$ ค่าน้ำหนักแบบถ่วงจะมีค่าเท่ากับ 0 ต่อมาทำการปรับเรียบข้อมูลอีกครั้งด้วยค่าน้ำหนักแบบถ่วง ค่าปรับเรียบข้อมูลสุดท้ายจะถูกคำนวณด้วยทั้งค่าน้ำหนักถ่วงเฉพาะส่วนและค่าน้ำหนักแบบถ่วง



รูปที่ 3.12 ขั้นตอนการปรับเรียบข้อมูลด้วยสมการถดถอยโดยพิจารณาเฉพาะส่วนแบบถ่วง

(ก) outlier มีผลกระทบต่อ การปรับเรียบข้อมูล

(ข) ค่าส่วนเหลือของ outlier มีค่ามากกว่า 6 เท่าของค่ามัธยฐานความผิดพลาดสัมบูรณ์ ดังนั้นค่าน้ำหนักแบบถ่วงมีค่าเท่ากับ 0

(ค) การปรับเรียบข้อมูลด้วยสมการถดถอยโดยพิจารณาเฉพาะส่วนแบบถ่วง

3.6 การวัดประสิทธิภาพ

3.6.1 การวัดประสิทธิภาพสำหรับการตรวจจับวัตถุ

ในงานวิจัยชิ้นนี้ การแสดงสัญลักษณ์ตำแหน่งเพื่อสื่อถึงพื้นที่ที่ได้จากการตรวจจับวัตถุ หรือข้อมูลตำแหน่งจริงของวัตถุ (Ground Truth : GT) จะถูกแสดงผ่านสัญลักษณ์กรอบสี่เหลี่ยม ล้อมรอบวัตถุนั้นๆ ซึ่งในงานวิจัยนี้วัตถุดังกล่าวคือคน

ในส่วนองวิธีการวัดประสิทธิภาพนั้นมีดังต่อไปนี้

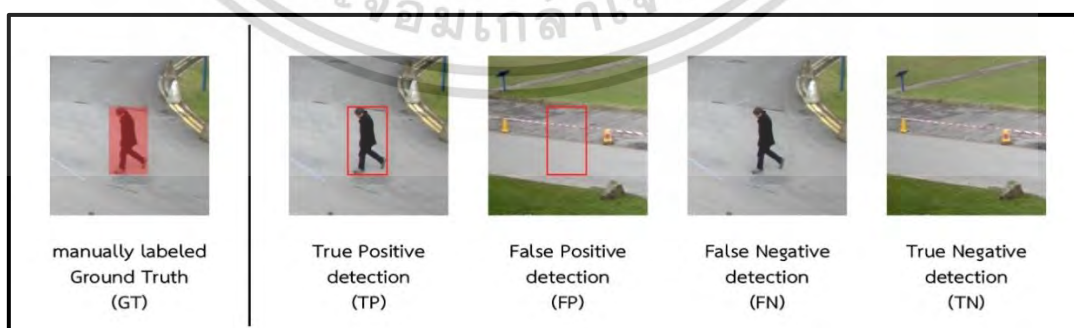
3.6.1.1 การวัดประสิทธิภาพความถูกต้องและความผิดพลาดในการตรวจจับ

ผลบวกจริง (True Positive : TP) , ผลบวกлож (False Positive : FP) ,
ผลลบจริง (True Negative : TN) , ผลลบлож (False Negative : FN)

ในการวัดประสิทธิภาพความถูกต้องในการตรวจจับ ในงานวิจัยส่วนใหญ่จะวัด ประสิทธิภาพในรูปของ ผลบวกจริง (True Positive : TP) ซึ่งเป็นกรณี พื้นที่ที่ตรวจจับได้มีวัตถุอยู่จริง และ ผลลบจริง (True Negative : TN) ซึ่งเป็นกรณี พื้นที่ที่ไม่มีการตรวจจับและไม่มีวัตถุอยู่จริง ส่วนการวัดประสิทธิภาพของการตรวจจับผิดพลาดในรูปของ ผลบวกлож (False Positive : FP) เป็นกรณี พื้นที่ที่ตรวจจับได้แต่ไม่มีวัตถุอยู่จริง ผลลบлож (False Negative : FN) เป็นกรณี พื้นที่ที่ไม่มีการตรวจจับแต่มีวัตถุอยู่จริง และ ซึ่งสามารถสรุปออกมาเป็น ตารางที่ 3.1 และรูปที่ 3.13

ตารางที่ 3.1 สรุปผลบวกจริง, ผลบวกлож, ผลลบจริง และ ผลลบлож

		ข้อมูลตำแหน่งจริง (Ground Truth : GT)	
		มีวัตถุ	ไม่มีวัตถุ
ผลการตรวจจับ	มีวัตถุ	ผลบวกจริง (TP)	ผลบวกлож (FP)
	ไม่มีวัตถุ	ผลลบлож (FN)	ผลลบจริง (TN)



รูปที่ 3.13 ข้อมูลตำแหน่งจริงและ

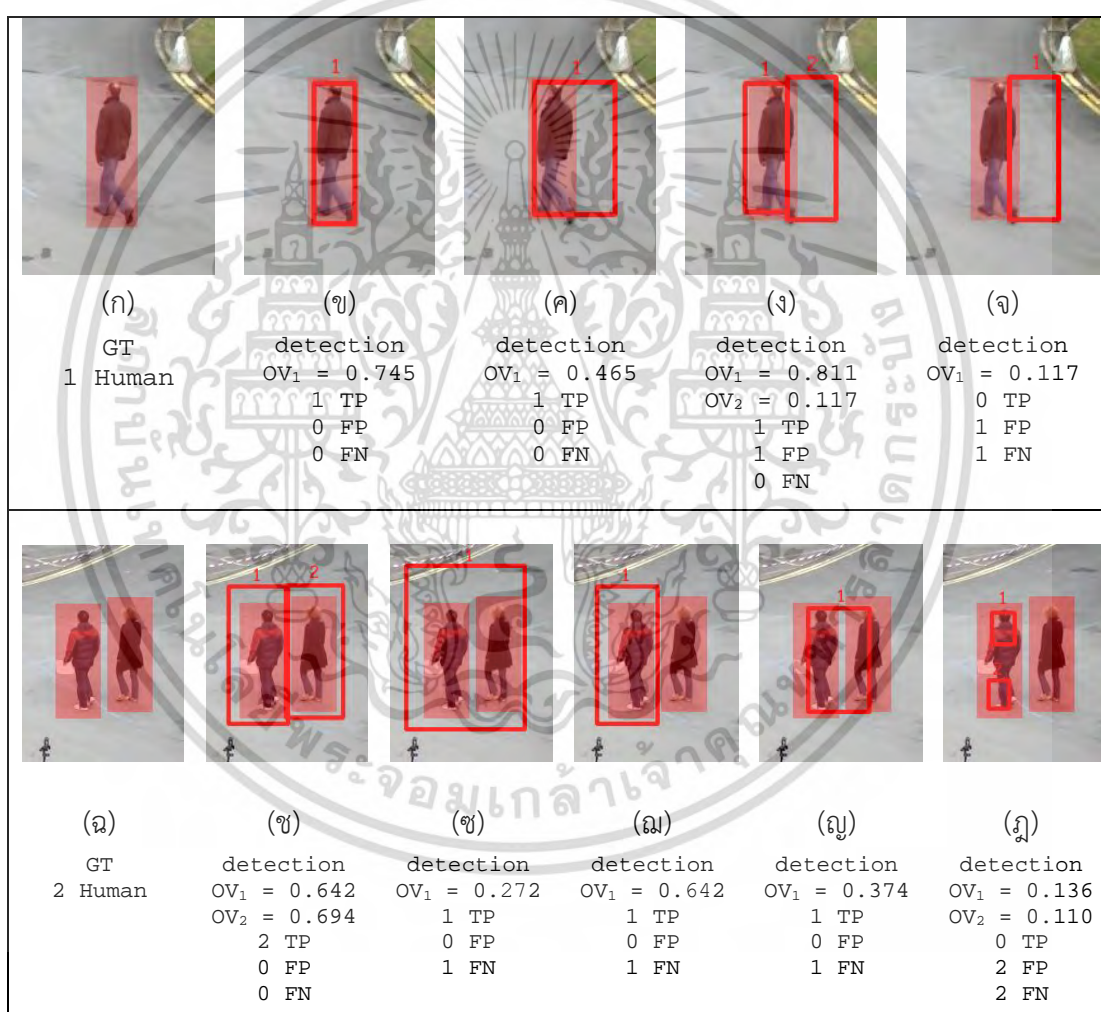
ผลลัพธ์การตรวจจับที่เป็นผลบวกจริง ผลบวกлож ผลลบложและผลลบจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากในงานวิจัยชิ้นนี้วัตถุจะถูกนำเสนอภายใต้สัญลักษณ์กรอบสี่เหลี่ยมล้อมรอบ ดังนั้นพื้นที่ที่ทับซ้อนกันระหว่างผลลัพธ์การตรวจจับวัตถุกับกรอบสี่เหลี่ยมของ GT จึงเป็นสิ่งสำคัญในการกำหนด TP, FP, FN ซึ่งในที่นี้กำหนดให้ T_{ov} คือ Overlap Threshold และการตรวจจับที่ D จะเป็น TP ต่อเมื่อ

$$OV = \frac{|D_i \cap G_i|}{|D_i \cup G_i|} \geq T_{ov} \quad (3.22)$$

โดย D_i และ G_i คือ คู่อันดับที่ i ระหว่างผลลัพธ์การตรวจจับและ GT



รูปที่ 3.14 การแยกแยะ TP,FP,FN ในกรณีต่างๆ โดยในตัวอย่างนี้กำหนดค่า $T_{ov} = 0.18$ เป็นเกณฑ์สำหรับการแยกแยะ

ในรูปที่ 3.14 จะเป็นการแยกแยะ TP , FP , FN ในกรณีต่างๆ ซึ่งได้กำหนดค่า $T_{ov} = 0.18$ สำหรับใช้ในการแยกแยะ โดยจากรูปที่ 3.14 พื้นที่ของ GT จะเป็นสี่เหลี่ยมที่มีพื้นที่เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภายในสีแดง และส่วนพื้นที่ผลการตรวจจับจะเป็นกรอบสีเหลี่ยมสีแดง โดยในรูปย่อยที่ (ข) - (จ) เป็นการแสดงถึงวิธีการนับ TP, FP, และ FN ที่มีความคลุมเคลือระหว่างการตรวจจับวัตถุกับความผิดพลาดที่ตรวจจับได้ส่วนของวัตถุ โดยผลการตรวจจับที่ 1 ในรูปย่อยที่ (ข) และผลการตรวจจับที่ 1 (OV_1) ในรูปย่อยที่ (ง) จะมีความแม่นยำ ที่ค่อนข้างสูงมากเพราะมีค่า OV มากเกินกว่า 0.7 ซึ่งถือว่า มีพื้นที่ส่วนทับซ้อนกันที่ค่อนข้างสูงและมีค่ามากกว่า T_{ov} จึงนับได้ว่าทั้งสองการตรวจจับนี้เป็น TP ส่วนผลการตรวจจับที่ 1 ในรูปย่อยที่ (ค) จะมีลักษณะที่ตรวจจับได้ส่วนของเงาวัตถุเข้ามาด้วย ซึ่งจะทำให้ค่า OV มีค่าน้อยลงไปจากผลการตรวจจับที่ 1 ในรูปย่อยที่ (ข) และ (ง) แต่ค่า OV ยังคงมีค่ามากกว่า T_{ov} จึงนับได้ว่าเป็น TP ส่วนผลการตรวจจับที่ 2 ในรูปย่อยที่ (ง) และผลการตรวจจับที่ 1 ในรูปย่อยที่ (จ) จะเป็นผลลัพธ์ที่ผิดพลาดที่ตรวจจับได้ส่วนของวัตถุมาเป็นผลลัพธ์และมีค่า OV ที่น้อยกว่า T_{ov} จึงนับได้ว่าเป็น FP และในรูปย่อยที่ (ฉ) จะนับว่ามี FN เพราะว่ามีพื้นที่ที่ตำแหน่งจริงมีวัตถุอยู่ แต่ผลการตรวจจับไม่ได้ให้ผลลัพธ์ในบริเวณนั้น ในรูปย่อยที่ (ช) - (ฎ) เป็นการแสดงถึงวิธีการนับ TP, FP, และ FN ที่มีผลลัพธ์การตรวจจับลักษณะที่รวมกันและแยกออกจากกัน โดยในรูปย่อยที่ (ช) จะเป็นผลลัพธ์การตรวจจับ 2 กรอบโดยตีกรอบล้อมรอบในแต่ละคน ซึ่งทั้งสองกรอบมีค่า OV มากกว่า T_{ov} จึงนับว่าเป็น TP ทั้ง 2 กรณี ในรูปย่อยที่ (ซ) จะมีผลลัพธ์การตรวจจับเพียง 1 แต่ GT มีอยู่ 2 คน จึงต้องเลือกคนที่มีค่า OV ที่มากกว่า OV อีกตัวหนึ่ง และค่า OV นั้นๆ ต้องมากกว่า T_{ov} จึงจะนำมานับเป็น 1 TP และอีกคนที่มีค่า OV น้อยกว่าจะนับเป็น 1 FN รูปย่อยที่ (ฌ) จะมีค่า 1 TP และ 1 FN รูปย่อยที่ (ญ) การตรวจจับจะเลือกจับคู่กับ GT ที่มีค่า OV มากกว่า และเนื่องจากค่า OV นั้นๆมากกว่า T_{ov} จึงนับเป็น 1 TP และอีก 1 GT ที่ไม่มีคู่การตรวจจับจึงนับเป็น 1 FN ส่วนในรูปย่อยที่ (ฎ) ทั้งสองการตรวจจับมีค่า OV น้อยกว่า T_{ov} จึงนับเป็น 2 FP และอีก 2 GT ที่เหลือนี้ไม่มีคู่การตรวจจับที่คู่ควรจึงนับกลายเป็น 2 FN ซึ่งในขั้นตอนนี้ก็จะทำให้ได้ค่า TP, FP, และ FN ของแต่ละเฟรมภาพแล้ว หลังจากนั้นจะนำค่าเหล่านี้ไปคำนวณเพื่อหาค่าต่างๆ ต่อไป

3.6.1.2 ค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) เป็นโอกาสของจำนวนคนที่ถูกตรวจจับได้อย่างถูกต้อง เทียบกับจำนวนคนที่ถูกตรวจจับได้ทั้งหมด ตามสมการที่ (3.23)

$$precision = \frac{TP}{TP + FP} \quad (3.23)$$

3.6.1.3 ค่าความครบถ้วน (Recall)

ค่าความครบถ้วน (Recall) เป็นโอกาสของจำนวนคนที่ถูกตรวจจับได้อย่างถูกต้อง เทียบกับจำนวนคนในข้อมูล ทั้งหมดจริง ตามสมการที่ (3.24)

$$recall = \frac{TP}{TP + FN} \quad (3.24)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษา อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.1.4 ค่าความถูกต้องในการตรวจจับหลายวัตถุ

(Multiple Object Detection Accuracy : MODA) [8]

เป็นการประเมินผลในแง่ความถูกต้องในการตรวจจับหลายวัตถุ ซึ่งจะใช้ค่าผลลบ
ลวงและผลบวกลวงมาใช้ในการคำนวณ โดยกำหนดให้ FN_t คือผลลบลวงในเฟรม t FP_t คือผลบวก
ลวงในเฟรม t TP_t คือ ผลบวกจริงในเฟรม t ซึ่งจะสามารถคำนวณค่าความถูกต้องในการตรวจจับ
หลายวัตถุ ตามสมการที่ 3.25

$$MODA(t) = 1 - \frac{c_m(FN_t) + c_f(FP_t)}{TP_t + FN_t} \quad (3.25)$$

โดยที่ c_m และ c_f เป็นค่าน้ำหนักสำหรับผลลบลวงและผลบวกลวงตามลำดับ ซึ่ง
งานวิจัยชิ้นนี้กำหนดให้ทั้งสองค่านี้นี้เท่ากับ 1 และจะสามารถคำนวณค่า Normalized MODA (N-
MODA) ตามสมการที่ 3.26

$$N - MODA = 1 - \frac{\sum_{t=1}^{N_{frames}} c_m(FN_t) + c_f(FP_t)}{\sum_{t=1}^{N_{frames}} TP_t + FN_t} \quad (3.26)$$

3.6.1.5 ค่าความแม่นยำในการตรวจจับหลายวัตถุ

(Multiple Object Detection Precision : MODP) [8]

จากข้อมูลการทับซ้อนระหว่างพื้นที่ตำแหน่งจริงของวัตถุกับผลลัพธ์ของการ
ตรวจจับที่ได้ สามารถนำมาคำนวณค่า Mapped Overlap Ratio ได้จากสมการที่ 3.27

$$Mapped\ Overlap\ Ratio = \sum_{i=1}^{N_{mapped}^{(t)}} \frac{|G_i^{(t)} \cap D_i^{(t)}|}{|G_i^{(t)} \cup D_i^{(t)}|} \quad (3.27)$$

โดยกำหนดให้ $G_i^{(t)}$ หมายถึง Ground truth ตัวที่ i ในเฟรมที่ t , $D_i^{(t)}$ หมายถึง
ผลลัพธ์การตรวจจับที่คู่กับ $G_i^{(t)}$ และ $N_{mapped}^{(t)}$ คือจำนวนของคู่ระหว่างการตรวจจับกับ Ground
truth ในเฟรมที่ t ซึ่งจะสามารถนำค่า Mapped Overlap Ratio มาทำการคำนวณค่าความแม่นยำ
ในการตรวจจับหลายวัตถุ (MODP) ในเฟรมที่ t โดยจะคำนวณได้จากสมการที่ 3.28

$$MODP(t) = \frac{(Mapped\ Overlap\ Ratio)}{N_{mapped}^{(t)}} \quad (3.28)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งในการคำนวณ $MODP(t)$ นั้น กำหนดให้ถ้าในเฟรมใดๆ มีค่า $N_{mapped}^{(t)} = 0$ จะกำหนดให้ค่า $MODP(t)$ เท่ากับ 0 จากนั้น เมื่อหาค่า $MODP(t)$ จากแต่ละเฟรมได้แล้ว จะสามารถนำมาหาค่า Normalized MODP (N-MODP) ต่อได้ในสมการที่ 3.29

$$N - MODP = \frac{\sum_{t=1}^{N_{frames}} MODP(t)}{N_{frames}} \quad (3.29)$$

3.6.2 การวัดประสิทธิภาพสำหรับการติดตามวัตถุ

3.6.2.1 ค่าความสมบูรณ์ของการติดตาม (Track Completeness : TC) [7]

ค่าความสมบูรณ์ของการติดตาม คำนวณจากจำนวนพิกัดที่คาบเกี่ยวกันระหว่างเส้นทางการเดินที่ถูกต้อง (Ground Truth) กับผลลัพธ์เส้นทางการเดินโดยพิจารณาเฟรมต่อเฟรมเทียบกับจำนวนพิกัดทั้งหมดของเส้นทางการเดินที่ถูกต้อง โดยยังมีค่ามากจะยิ่งถือว่าผลลัพธ์เส้นทางการเดินมีตำแหน่งที่ใกล้กับเส้นทางการเดินที่ถูกต้องและมีประสิทธิภาพการติดตามที่ดี โดยสามารถคำนวณค่าความสมบูรณ์ของการติดตามจากสมการที่ 3.30

$$TC = \frac{\sum_{k=1}^N O(GT_{ik}, ST_{jk})}{\text{number of } GT_i} \quad (3.30)$$

$$O(GT_{ik}, ST_{jk}) = \begin{cases} 1 & \text{if } dist(GT_{ik}, ST_{jk}) < T_{dist} \\ 0 & \text{if } dist(GT_{ik}, ST_{jk}) \geq T_{dist} \end{cases} \quad (3.31)$$

กำหนดให้ GT_{ik} คือ ตำแหน่งของคำตอบเส้นทางการเดินที่ถูกต้องในเฟรมที่ k ST_{jk} คือ ตำแหน่งของคำตอบเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในเฟรมที่ k N คือ จำนวนเฟรมของ $GT_i \cap ST_j$ และ T_{dist} คือ ค่าขีดแบ่งระยะห่างระหว่างจุด

3.6.2.2 ค่าระยะที่ผิดพลาดของการติดตาม (Tracking Matching Error : TME) [7]

ค่าระยะที่ผิดพลาดของการติดตาม คือค่าเฉลี่ยของระยะห่างระหว่างจุดคำตอบเส้นทางการเดินที่ถูกต้อง และเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในแต่ละเฟรม โดยค่านี้นั้นยังมีค่าน้อยหมายความว่าเส้นทางการเดินที่ทดสอบมีประสิทธิภาพในด้านของความแม่นยำ โดยสามารถคำนวณค่าระยะที่ผิดพลาดของการติดตามได้จากสมการที่ 3.32

$$TME = \frac{\sum_{k=1}^N \sqrt{(X_{GT_{ik}} - X_{ST_{jk}})^2 + (Y_{GT_{ik}} - Y_{ST_{jk}})^2}}{N} \quad (3.32)$$

จากสมการที่ 3.32 กำหนดให้ $X_{GT_{ik}}$ คือ พิกัดแกน x ของจุดคำตอบเส้นทางการเดินที่ถูกต้องในเฟรมที่ k $X_{ST_{jk}}$ คือ พิกัดแกน x ของจุดคำตอบเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในเฟรมที่ k $Y_{GT_{ik}}$ คือ พิกัดแกน y ของจุดคำตอบเส้นทางการเดินที่ถูกต้องในเฟรมที่ k $Y_{ST_{jk}}$ คือ พิกัดแกน y ของจุดคำตอบเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในเฟรมที่ k N คือ จำนวนเฟรมของ GT_i \cap ST_j

3.6.2.3 การวัดระยะแบบยูคลิเดียน (Euclidean Distance)

การวัดระยะแบบยูคลิเดียน เป็นการวัดค่าระยะห่างระหว่างเส้นทางการเดินที่ถูกต้อง กับเส้นทางการเดินที่ต้องการวัดประสิทธิภาพ เฟรม โดยค่านี้นั้นยังมีค่าน้อยหมายความว่าเส้นทางการเดินที่ทดสอบมีประสิทธิภาพในด้านของความแม่นยำ สามารถคำนวณการวัดระยะแบบยูคลิเดียนได้จากสมการที่ 3.33

$$Euclidean\ Distance = \sqrt{\sum_{k=1}^N (X_{GT_{ik}} - X_{ST_{jk}})^2 + (Y_{GT_{ik}} - Y_{ST_{jk}})^2} \quad (3.33)$$

จากสมการที่ 3.33 กำหนดให้ $X_{GT_{ik}}$ คือ พิกัดแกน x ของจุดคำตอบเส้นทางการเดินที่ถูกต้องในเฟรมที่ k $X_{ST_{jk}}$ คือ พิกัดแกน x ของจุดคำตอบเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในเฟรมที่ k $Y_{GT_{ik}}$ คือ พิกัดแกน y ของจุดคำตอบเส้นทางการเดินที่ถูกต้องในเฟรมที่ k $Y_{ST_{jk}}$ คือ พิกัดแกน y ของจุดคำตอบเส้นทางการเดินที่ต้องการวัดประสิทธิภาพในเฟรมที่ k N คือ จำนวนเฟรมของ GT_i \cap ST_j

3.6.2.4 ค่าความถูกต้องในการติดตามหลายวัตถุ (MOTA) [8][9]

เป็นการวัดค่าความถูกต้องของผลลัพธ์การติดตามวัตถุ โดยพิจารณาจาก จำนวนวัตถุที่ติดตามพลาด ค่าผลบวกหลง และการติดตามสลับบุคคลของผลลัพธ์การติดตามวัตถุเทียบกับการติดตามวัตถุที่ถูกต้อง โดยสามารถวัดค่าความถูกต้องในการติดตามหลายวัตถุจากสมการที่ 3.34

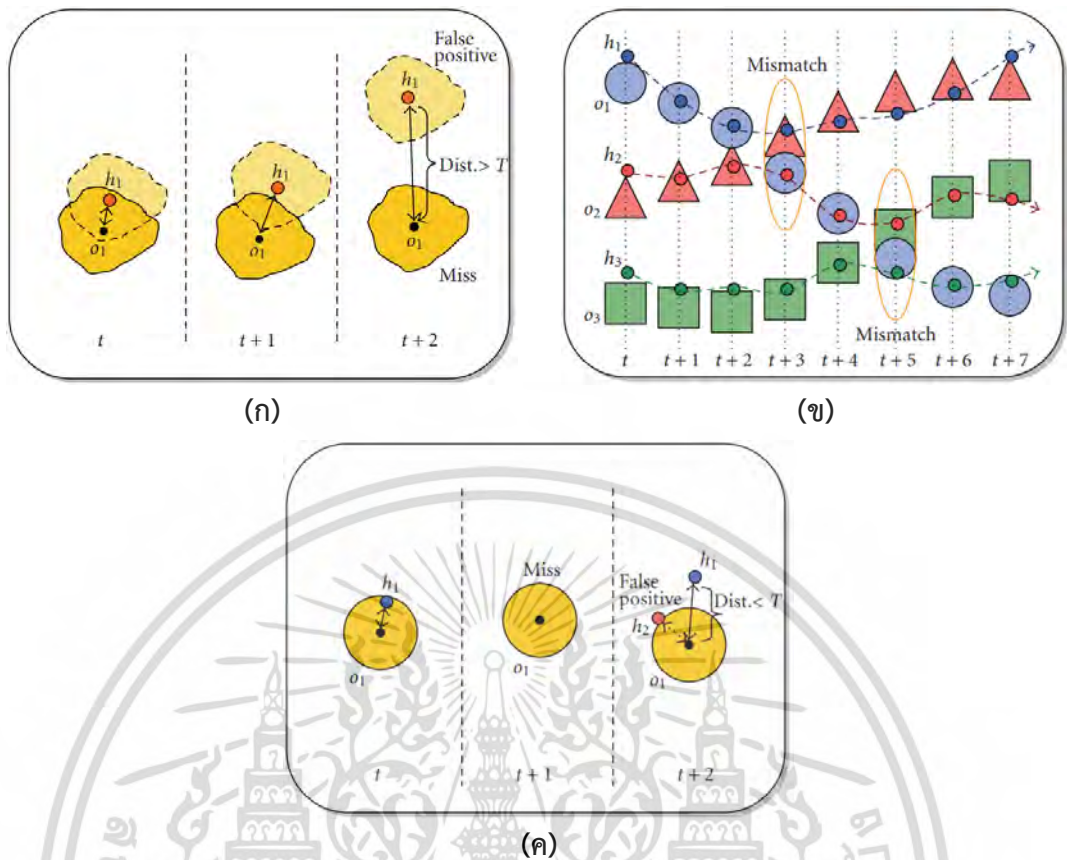
$$MOTA = 1 - \frac{\sum_{t=1}^{N_{frames}} ((m_t) + (fp_t) + (ID - SWITCHES_t))}{\sum_{t=1}^{N_{frames}} N_G^{(t)}} \quad (3.34)$$

กำหนดให้ m_t คือจำนวนวัตถุที่ติดตามพลาดในเฟรม t fp_t คือจำนวนผลบวกหลงในเฟรม t และ $ID - SWITCHES_t$ คือจำนวนของการติดตามตัวสลับบุคคลที่เฟรม t โดยจะพิจารณาเทียบกับติดตามบุคคลในเฟรม $(t - 1)$ และ $N_G^{(t)}$ คือ จำนวน ground truth ในเฟรม t

การจะวัดค่าความถูกต้องในการติดตามหลายวัตถุจะต้องทำการจับคู่ระหว่างผลการติดตามกับ ground truth เพื่อเป็นการระบุว่าผลการติดตามวัตถุดังกล่าวหมายถึงวัตถุชิ้นใด เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปเผยแพร่บนสื่อออนไลน์ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้ $\{h_1, h_2, \dots, h_m\}$ หมายถึง กลุ่มผลการติดตามในแต่ละเฟรม $\{o_1, o_2, \dots, o_n\}$ หมายถึง กลุ่ม ground truth ของวัตถุในแต่ละเฟรม $M_t = \{(o_i, h_j)\}$ คือผลการจับคู่ที่เหมาะสมระหว่าง ground truth และผลการติดตามวัตถุในเฟรม t โดยในตอนเริ่มต้นการวัดกำหนดให้ $M_0 = \{\cdot\}$ สำหรับขั้นตอนการจับคู่ ground truth กับผลการติดตามเพื่อวัดประสิทธิภาพ การนับจำนวนวัตถุที่ติดตามพลาด จำนวนผลบวกของจำนวนของการติดตามตัวสลับบุคคล ในแต่ละเฟรม มีขั้นตอนและเงื่อนไขดังต่อไปนี้

- 1) ในเฟรมที่ $t-1$ มีการจับคู่ (o_i, h_j) ของ M_{t-1} ถ้าหากในเฟรมที่ t วัตถุ o_i และผล h_j ยังคงปรากฏ และมีระยะห่างระหว่างกันน้อยกว่าค่าขีดแบ่ง T จะนับว่าการจับคู่ยังคงเหมือนเดิม การจับคู่ (o_i, h_j) จะยังคงอยู่ในเฟรมที่ t
- 2) สำหรับวัตถุที่ไม่มีข้อมูลการจับคู่ในเฟรมก่อนหน้ามาก่อน การจะหาคู่ที่เหมาะสมจะจับคู่แบบ one-to-one เท่านั้น และระยะห่างของวัตถุกับผลการติดตามที่จะจับคู่ต้องน้อยกว่าค่าขีดแบ่ง T ซึ่งการจับคู่จะต้องจับคู่อย่างเหมาะสมด้วยเกณฑ์ที่ผลรวมระยะห่างระหว่างแต่ละวัตถุกับผลการติดตามจะต้องมีค่าน้อยที่สุด ซึ่งปัญหาดังกล่าวคือปัญหาการมอบหมายงาน (Assignment Problem) แบบค่าน้ำหนักรวมน้อยที่สุด และสามารถแก้ปัญหาได้ด้วยอัลกอริทึม Munkres ถ้าเกิดกรณีในเฟรม t คู่ (o_i, h_k) นั้นไม่ตรงกับคู่ (o_i, h_j) ใน M_{t-1} คู่ (o_i, h_j) จะถูกแทนที่ด้วย (o_i, h_k) ใน M_t ซึ่งค่า $ID - SWITCHES_t$ จะถูกนับขึ้นในเฟรม t
- 3) หลังจากทำขั้นตอนที่ 1 และ 2 แล้ว จะได้กลุ่มการจับคู่ที่เหมาะสมในแต่ละเฟรม กำหนดให้ N_{mapped}^t คือจำนวนคู่ที่เหมาะสมในเฟรม t และ d_t^i คือระยะห่างระหว่างวัตถุ o_i กับผลการติดตามที่ถูกจับคู่ด้วย โดยทั้งสองค่านี้จะถูกนำไปใช้ในการคำนวณหาความแม่นยำในการติดตามหลายวัตถุ (MOTP) ซึ่งจะอธิบายในหัวข้อถัดไป
- 4) สำหรับผลการติดตามที่ไม่ถูกจับคู่ที่เหลือจะนับว่าเป็น fp_t และสำหรับวัตถุที่ไม่ถูกจับคู่ที่เหลือจับคู่กันเป็นค่า m_t สำหรับเฟรม t
- 5) ทำขั้นตอนที่ 1 ถึง 4 ซ้ำอีกครั้งในเฟรมถัดไป โดยในตอนเริ่มต้นการวัดในเฟรมแรก จากเงื่อนไขกำหนดให้ค่าตั้งต้น $M_0 = \{\cdot\}$ ซึ่งหมายความว่าจะไม่มีการนับค่า $ID - SWITCHES_t$ เกิดขึ้นในเฟรมเริ่มต้น



รูปที่ 3.15 การนับจำนวนวัตถุที่ติดตามพลาด จำนวนผลบวกกลาง
จำนวนของการติดตามตัวสลับบุคคล ในกรณีต่างๆ

(ก) การนับจำนวนวัตถุที่ติดตามพลาด จำนวนผลบวกกลาง

(ข) การนับจำนวนของการติดตามสลับตัวบุคคล

(ค) การเลือกจับคู่ในกรณีที่มีผลการติดตามหลายผล

จากรูปที่ 3.15 (ก) เป็นตัวอย่าง เมื่อระยะห่างระหว่างวัตถุ o_1 และผล h_1 มีค่ามากกว่าค่าขีดแบ่ง T วัตถุทั้งสองจะถือว่าไม่มีความสัมพันธ์กัน ดังนั้นวัตถุ o_1 จะนับว่าเป็นค่า m_{t+2} และผล h_1 จะนับว่าเป็นค่า fp_{t+2} รูปที่ 3.15 (ข) เป็นตัวอย่างกรณีที่เกิดการติดตามสลับตัว โดยที่ในตอนแรกผล h_2 จะจับคู่ว่าติดตามวัตถุ o_2 อยู่ หลังจากนั้นในเฟรมที่ $t+3$ วัตถุ o_1 และวัตถุ o_2 เกิดการเดินที่มีลักษณะไขว่กัน ทำให้ผล h_2 เกิดการติดตามผิดวัตถุเกิดขึ้น และจากตัวอย่างรูปที่ 3.15 มีการติดตามสลับวัตถุเกิดขึ้นอีกครั้งในเฟรมที่ $t+5$ ที่วัตถุ o_3 รูปที่ 3.15 (ค) เป็นกรณีการจับคู่การติดตามใหม่หลังจากผลการติดตามหายไป เฟรม t วัตถุ o_1 จะถูกติดตามด้วยผล h_1 ต่อมาในเฟรม $t+1$ ผลลัพธ์การติดตามสูญหายไป ในเฟรม $t+2$ มีผลการติดตามอยู่ 2 คาตอบด้วยกัน โดยการจับคู่จะเลือกจับคู่วัตถุ o_1 กับผล h_1 ที่ถึงแม้ผล h_2 จะอยู่ใกล้กับวัตถุ o_1 มากกว่าก็ตาม เนื่องจากการจับคู่จะอ้างอิงกับการจับคู่ที่เกิดขึ้นล่าสุดที่ได้เคยจับคู่กันไว้ ส่วนผล h_2 จะถูกนับว่าเป็นผลที่เป็น fp_t

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.2.5 ค่าความแม่นยำในการติดตามหลายวัตถุ (MOTP) [8]

ค่าความแม่นยำ ในการติดตามหลายวัตถุ คือ การรวมผลความคลาดเคลื่อนระหว่างวัตถุและ ground truth ที่เป็นคู่วัตถุที่เหมาะสมกัน แล้วทำการเฉลี่ยค่าด้วยจำนวนคู่ทั้งหมด เพื่อประเมินความสามารถในแง่ของความแม่นยำ ของวิธีการที่ใช้ในการติดตามวัตถุ โดยสามารถคำนวณค่าความแม่นยำ ในการติดตามหลายวัตถุได้จากสมการที่ 3.35

$$MOTP = \frac{\sum_{i=1}^{N_{mapped}} \sum_{t=1}^{N_{frames}^{(t)}} d_t^i}{\sum_{t=1}^{N_{frames}} N_{mapped}^{(t)}} \quad (3.35)$$

กำหนดให้ d_t^i หมายถึงระยะห่างระหว่าง ground truth กับผลการติดตามของคู่ที่เหมาะสมกัน N_{mapped} หมายถึงจำนวนคู่วัตถุที่เหมาะสมกันของผลการติดตามกับ ground truth และ N_{mapped}^t หมายถึงจำนวนคู่วัตถุที่เหมาะสมกันในเฟรม t

3.7 ค่าเฉลี่ยเรขาคณิต (Geometric mean)

เป็นค่าเฉลี่ยชนิดหนึ่ง นิยมใช้กับการหาค่ากลางของข้อมูลที่เกิดจากการถดถอยที่ N จากข้อมูล N ชุด โดยคำนวณได้จากสมการที่ 3.36

$$G.M. = \sqrt[n]{x_1 x_2 x_3 \dots x_n} \quad (3.36)$$

กำหนดให้ x คือข้อมูลที่ต้องการจะหาค่าเฉลี่ยเรขาคณิต และ n คือจำนวนข้อมูล

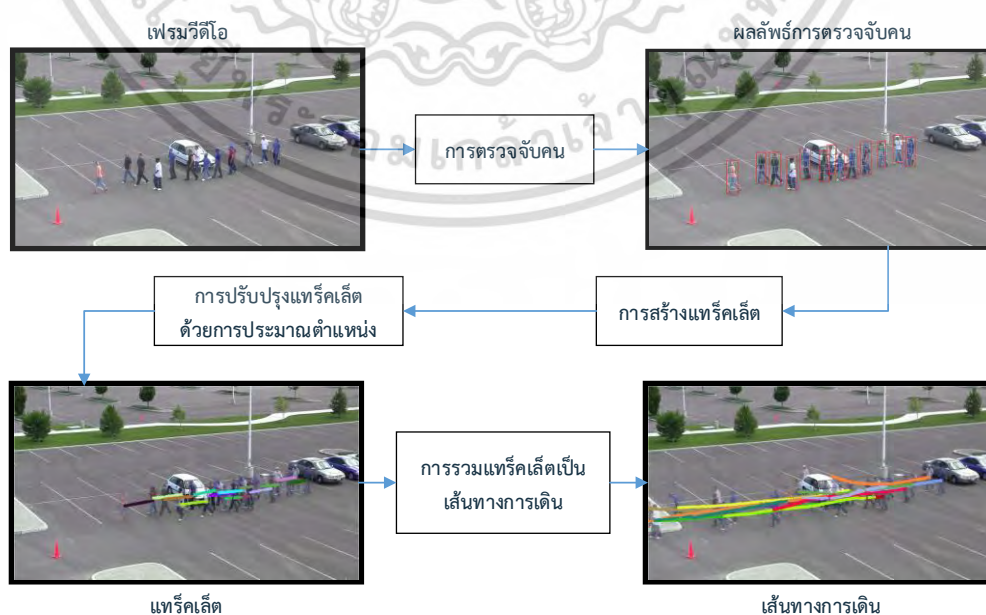
บทที่ 4

งานวิจัยที่นำเสนอ

การติดตามวัตถุ (Object Tracking) นั้นเป็นงานวิจัยที่สำคัญยิ่งในวงการของคอมพิวเตอร์วิทัศน์ โดยเป้าหมายหลักของการทำการติดตามวัตถุคือการสร้างเส้นทางการเดินของวัตถุ จากตำแหน่งวัตถุของแต่ละเฟรมของวิดีโอ บทนี้จะกล่าวถึงการเชื่อมโยงข้อมูลคนๆเดียวกันเพื่อหาเส้นทางการเดิน จากการตรวจจับคนในแต่ละเฟรมของวิดีโอ แนวคิดที่ใช้คือ คือ ตรวจจับบุคคลจากแต่ละเฟรมวิดีโอ แล้วทำการแบ่งหาเส้นทางการเดินช่วงสั้นๆจากช่วงส่วนย่อยวิดีโอก่อน แล้วจึงค่อยทำการเชื่อมโยงข้อมูลในการรวมเส้นทางการเดินช่วงสั้นๆนี้ให้กลายเป็นเส้นทางการเดินตลอดทั้งวิดีโอ ในงานวิจัยที่นำเสนอนี้เรียกเส้นทางการเดินช่วงสั้นๆว่า แทร็คเล็ต (Tracklet) การแบ่งส่วนย่อยวิดีโอเกณฑ์ในการแบ่งจะแบ่งที่แต่ละจำนวนเฟรมที่คนเคลื่อนไหวด้วยความเร็วคงที่และเป็นเส้นตรง และจะสร้างแทร็คเล็ตโดยพิจารณาภายในแต่ละส่วนย่อยของวิดีโอ ดังนั้นลักษณะของแทร็คเล็ตก็จะเป็นเส้นทางการเดินช่วงสั้นๆของคนที่มีความเร็วคงที่และเป็นเส้นตรงด้วยเช่นกัน จากนั้นจึงรวมแต่ละแทร็คเล็ตเป็นเส้นทางการเดินของแต่ละคนตลอดทั้งวิดีโอ

4.1 ภาพรวมงานวิจัยที่นำเสนอ

งานวิจัยที่นำเสนอ สามารถแบ่งเป็นขั้นตอนหลักๆ ได้ 4 ขั้นตอน 1.การตรวจจับคน 2.การสร้างแทร็คเล็ต 3.การปรับปรุงแทร็คเล็ตด้วยการประมาณตำแหน่ง 4.การรวมแทร็คเล็ตเป็นเส้นทางการเดิน



รูปที่ 4.1 ภาพรวมงานวิจัยที่นำเสนอ

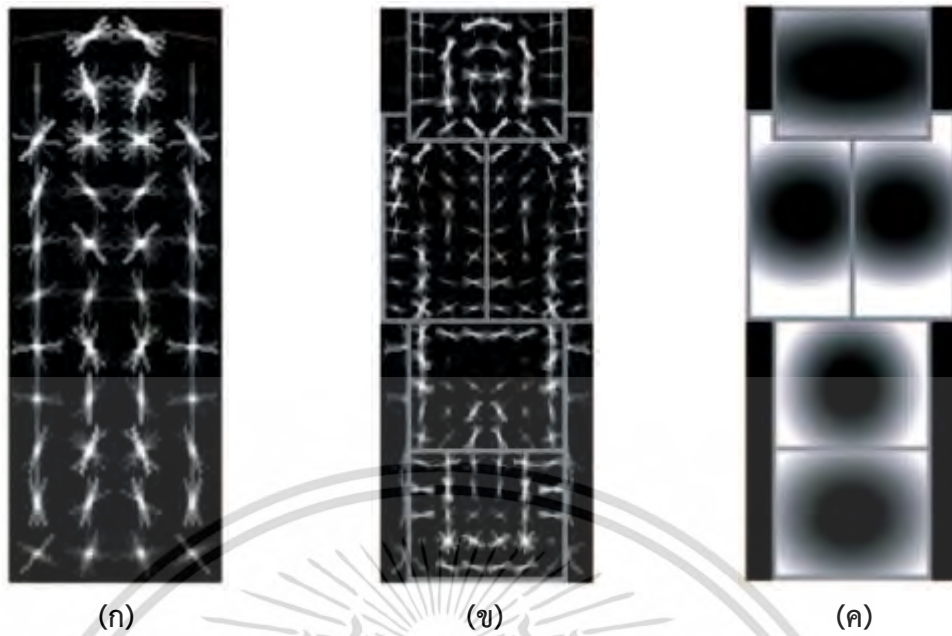
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.1 เป็นภาพรวมขั้นตอนทั้งหมดของงานวิจัยที่นำเสนอ โดยในขั้นตอนที่ 1 ภาพนิ่งของเฟรมวีดีโอจะถูกนำมาตรวจจับคน ด้วยวิธีการตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน[1] ซึ่งจะทำให้ได้ผลลัพธ์การตรวจจับเป็นตำแหน่งล้อมรอบภาพคนเต็มตัว (bounding box) และตำแหน่งล้อมรอบส่วนแต่ละส่วนของคน ขั้นตอนที่ 2 เป็นการนำผลลัพธ์ของการตรวจจับคนมาทำการเชื่อมโยงข้อมูลเพื่อหาเส้นทางการเดินสั้นๆ ซึ่งในงานวิจัยที่เสนอนี้เรียกว่าแทร็คเล็ต (Tracklet) โดยแต่ละแทร็คเล็ตนั้นเป็นผลเส้นทางการเดินสั้นๆที่ได้จากการเชื่อมโยงข้อมูลภายในช่วงเฟรมที่แบ่งจำนวนเฟรมเป็นช่วงสั้นๆภายใต้สมมุติฐานที่ว่า เป็นช่วงที่คนเดินด้วยความเร็วคงที่และเดินเป็นเส้นตรง สลับกันในขั้นตอนถัดมา ผลลัพธ์แทร็คเล็ตที่ได้มีโอกาสที่จะมีการเชื่อมโยงข้อมูลที่ผิดพลาด ซึ่งมีสาเหตุมาจาก ผลการตรวจจับที่ผิดพลาด หรือคนเดินถูกบดบัง เป็นต้น ส่งผลทำให้แทร็คเล็ตมีการเชื่อมโยงสลับคนเกิดขึ้นได้ จึงต้องนำมาทำการปรับปรุงในขั้นตอนที่ 3 ซึ่งเป็นขั้นตอนการปรับปรุงแทร็คเล็ตด้วยการประมาณตำแหน่ง เพื่อให้แทร็คเล็ตมีประสิทธิภาพที่ดีขึ้น ต่อมาเป็นระดับขั้นตอนในการเชื่อมโยงข้อมูลแทร็คเล็ตจากหลายๆกลุ่ม เพื่อสร้างเส้นทางการเดินที่สมบูรณ์ตลอดทั้งวีดีโอของบุคคลแต่ละคน

4.2 การตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน (Object Detection with Discriminatively Trained Part-Based Models) [1]

เทคนิคที่ใช้ในการตรวจจับภาพคนมีหลากหลายงานวิจัยที่ได้คิดค้นและนำเสนอ หนึ่งในวิธีการที่นิยมคือ งานวิจัย[10] ได้นำเสนอการตรวจจับภาพคนโดยจุดสำคัญคือการใช้ค่าความถี่ของทิศเกรเดียนท์ (Histograms of Oriented Gradients : HOG) พร้อมด้วยการใช้ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) [10] ในการจัดกลุ่มคุณลักษณะของความถี่ของทิศเกรเดียนท์ว่าเวกเตอร์คุณลักษณะนั้นๆ เป็นรูปคนหรือไม่ และยังใช้เทคนิคอิมเมจพีระมิด (Image Pyramid) เพื่อที่จะทำการตรวจจับภาพหลากหลายขนาด ต่อมางานวิจัย[1] ได้นำเสนอแนวคิดตัวกรอง root (root filter) ตัวกรองส่วน (Part filter) และ โมเดลส่วนที่ปรับรูปได้ (deformable part model) โดยใช้ค่าความถี่ของทิศเกรเดียนท์เป็นคุณลักษณะ ตัวกรอง root คือเวกเตอร์น้ำหนักของวัตถุทั้งชิ้น ตัวกรองส่วนคือเวกเตอร์น้ำหนักของแต่ละส่วนของวัตถุ ยกตัวอย่างเช่น หากวัตถุเป็นคน แต่ละส่วนของวัตถุจะเป็น ส่วนหัว ส่วนล ตัวส่วนแขน ส่วนขา เป็นต้น โดยตัวกรองส่วนจะมีความละเอียดที่สูงกว่าตัวกรอง root ในการตรวจจับภาพคนยังคำนวนที่ภาพหลากหลายขนาดด้วยการใช้เทคนิคอิมเมจพีระมิด และใช้เทคนิค Latent SVM ในการฝึกสอนโมเดล

สำหรับขั้นตอนของการตรวจจับบุคคลนี้งานวิจัยที่นำเสนอได้ใช้วิธีการจากงานวิจัย[1] ในการตรวจจับคนและแบ่งส่วนย่อยรูปคน เพื่อจะนำไปเชื่อมโยงข้อมูลเพื่อติดตามบุคคลต่อไป



รูปที่ 4.2 โมเดลประเภทคน

(ก) ตัวกรอง Root

(ข) ตัวกรองส่วน

(ค) โมเดลส่วนที่ปรับรูปได้

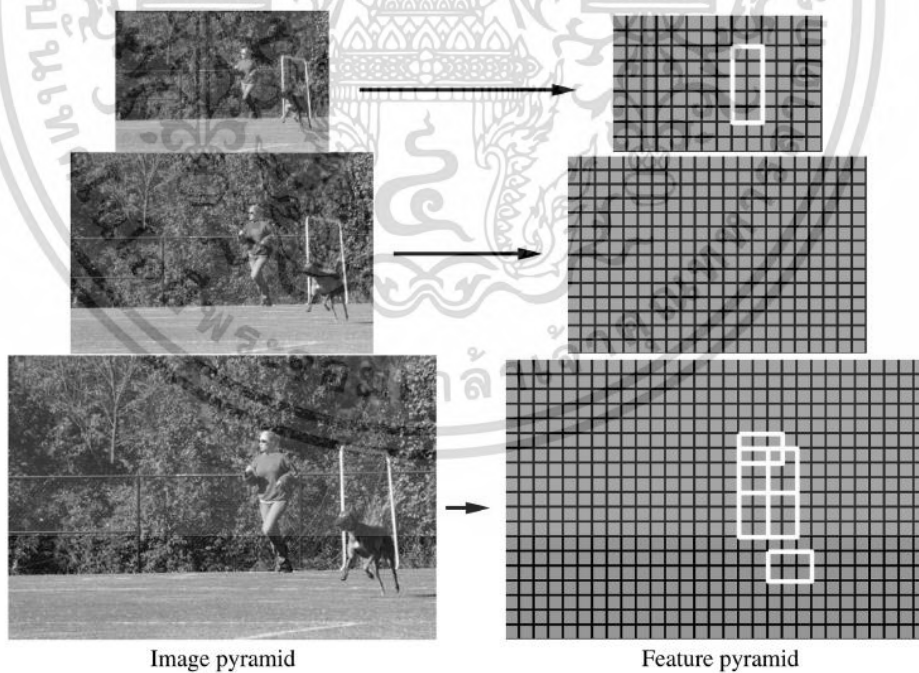


Image pyramid

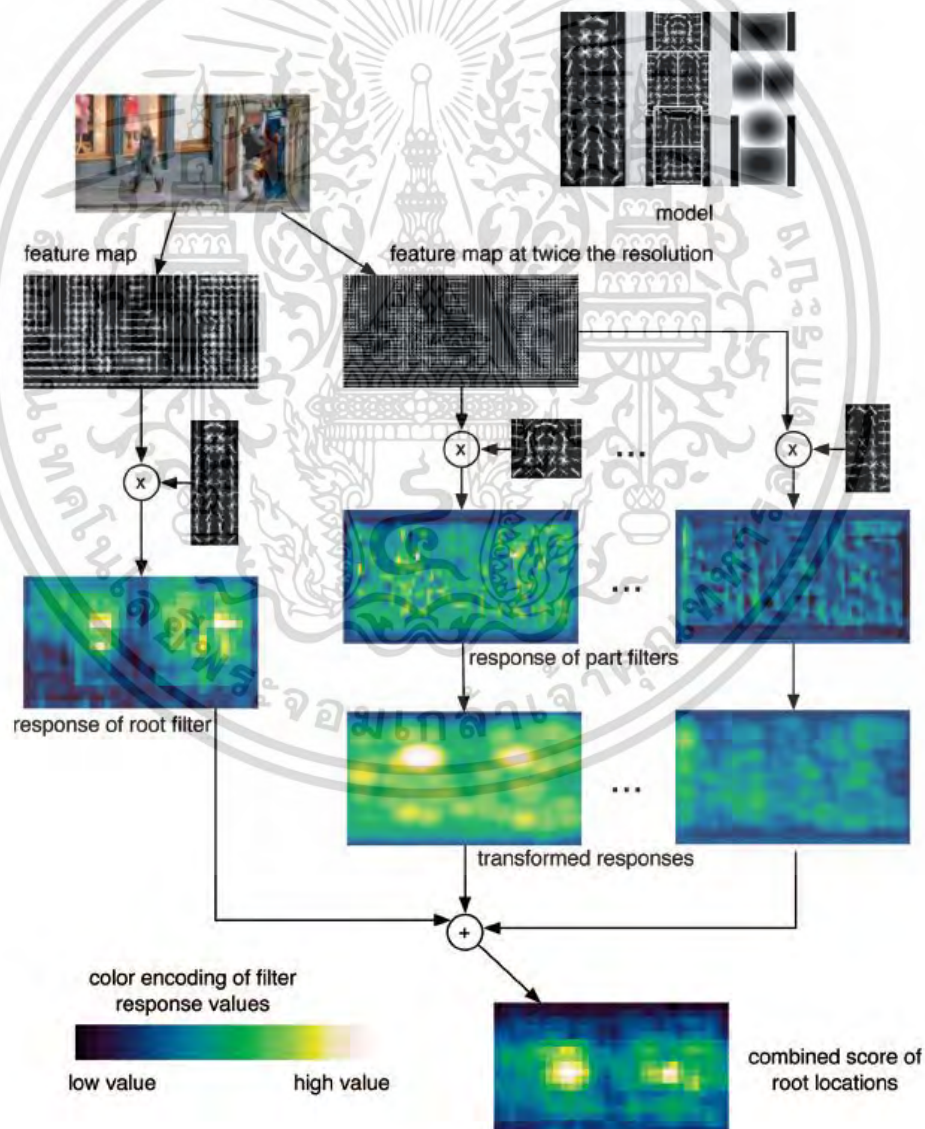
Feature pyramid

รูปที่ 4.3 อิมเมจพีระมิตและคุณลักษณะพีระมิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1 ขั้นตอนการตรวจจับ

เริ่มจากการท ขิมเมจพีระมิดกับรูปภาพที่ต้องการจะท การตรวจจับ โดยในภาพ 1 ขนาด จะผ่านการแปลงให้อยู่ในรูปแบบคุณลักษณะค่าความถี่ของทศเกรเดียนท์ โดยความละเอียดของภาพ ที่ใช้สำหรับคานวณของวัตถุแต่ละส่วนจะต้องมีความละเอียดเป็นเท่าตัวของขนาดที่ใช้คานวณแบบ เต็มวัตถุ ต่อมาใช้ตัวกรอง root และตัวกรองส่วนแยกกันคานวณที่ความละเอียดที่ต่างกัน นาผลลัพธ์ การกรองที่ได้จากตัวกรองส่วน ไปท การแปลงค่าผลลัพธ์เพื่อให้กระจายค่าคะแนนการกรองที่สูงไปยัง ตำแหน่งใกล้เคียง ต่อมานำค่าผลลัพธ์การกรองที่ได้จากตัวกรอง root มารวมคะแนนกับค่าผลลัพธ์ที่ ผ่านการแปลงของแต่ละส่วน ผลลัพธ์ที่ได้จากการรวมคะแนน ตำแหน่งที่มีคะแนนสูงมากกว่าค่าขีด แบ่งจะท การเลือกท าแห่งนั้นเป็นท าแห่งยึดของผลลัพธ์การตรวจจับ



รูปที่ 4.4 ภาพรวมขั้นตอนการตรวจจับของงานวิจัย[1]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้ $(F_0, P_1, \dots, P_n, b)$ คือโมเดลของวัตถุที่แบ่งส่วน n ส่วน โดยให้ F_0 คือตัวกรอง root P_i คือ โมเดลส่วนที่ i และ b คือค่า bias (F_i, v_i, d_i) คือโมเดลส่วนที่ i โดย F_i คือตัวกรองของส่วนที่ i v_i คือเวกเตอร์ 2 มิติของค่าระยะห่างระหว่างตำแหน่งยึดของตัวกรอง root และตัวกรองส่วนที่ i d_i คือเวกเตอร์ 4 มิติที่ระบุค่าสัมประสิทธิ์ของฟังก์ชันพหุนามดีกรีสอง (Quadratic Function) ที่กำหนดค่าการปรับรูปได้ (deformation cost) สำหรับแต่ละตำแหน่งที่เป็นไปได้ของส่วนที่สัมพันธ์กับตำแหน่งยึด

$$z = (p_0, \dots, p_n) \quad (4.1)$$

$$p_i = (x_i, y_i, l_i) \quad (4.2)$$

กำหนดให้ p_i คือตำแหน่งของตัวกรองส่วนที่ i อยู่ที่ตำแหน่ง (x, y) ในระดับชั้นที่ l สำหรับระดับชั้นที่คำนวณตัวกรองส่วนจะต้องมีความละเอียดเป็นสองเท่าของความละเอียดที่คำนวณตัวกรอง root

$$R_{i,l}(x, y) = F'_i \cdot \phi(H, (x, y, l)) \quad (4.3)$$

กำหนดให้ F'_i คือเวกเตอร์นำหน้าของตัวกรอง F_i ที่เรียงลำดับแบบแถวเป็นหลัก H คือคุณลักษณะพีระมิด $\phi(H, (x, y, l))$ คือเวกเตอร์ของคุณลักษณะหน้าที่ย่อยของ H มีตำแหน่งซ้ายบนที่ตำแหน่ง (x, y) ในชั้นที่ l ที่เรียงลำดับแบบแถวเป็นหลัก (Row-major order) $R_{i,l}(x, y)$ เป็นอาร์เรย์ที่เก็บผลคำตอบของผลลัพธ์การกรองคุณลักษณะพีระมิดจากตัวกรองส่วนที่ i ในชั้นที่ l โดยเป็นผลจากการหาความสัมพันธ์ (cross correlation) ระหว่างตัวกรองส่วน F_i กับคุณลักษณะพีระมิดชั้นที่ l หลังจากคำนวณผลการกรอง จะทำการแปลงคำตอบของการกรองจากแต่ละส่วน เพื่อให้เกิดความไม่แน่นอนเชิงตำแหน่ง

$$D_{i,l}(x, y) = \max_{dx, dy} (R_{i,l}(x + dx, y + dy) - d_i \cdot \phi(dx, dy)) \quad (4.4)$$

จากสมการ 4.4 การแปลงจะกระจายค่าคะแนนการกรองที่สูงไปยังตำแหน่งใกล้เคียง $D_{i,l}(x, y)$ คือ ค่าสนับสนุนสูงสุดของส่วนที่ i ที่จะนำไปใช้ในการคำนวณคะแนนของตำแหน่ง root ที่ตำแหน่ง (x, y) ในชั้นที่ l โดยอาร์เรย์ $D_{i,l}$ สามารถแปลงจากอาร์เรย์ $R_{i,l}$ โดยใช้วิธีการแปลงเชิงระยะห่าง (the generalized distance transform algorithm) จากงานวิจัย [11]

$$\text{score}(x_0, y_0, l_0) = R_{0,l_0}(x_0, y_0) + \sum_{i=1}^n D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b \quad (4.5)$$

กำหนดให้ $l_i = l_0 - \lambda$ โดยที่ $i > 0$ ซึ่ง λ เป็นระยะห่างจำนวนของชั้นคุณลักษณะพีระมิดระหว่างชั้นที่ใช้เป็นคุณลักษณะเพื่อใช้ในการหาผลลัพธ์การกรองของตัวกรอง root และชั้นที่ใช้เป็นคุณลักษณะเพื่อใช้ในการหาผลลัพธ์การกรองของตัวกรองส่วน จากสมการที่ 4.5 คะแนนรวมของ root ที่ระดับชั้น l_0 คิดจาก ผลรวมของผลลัพธ์การกรองจากตัวกรอง root ที่ระดับชั้นดังกล่าวบวกด้วยผลจากการแปลงค่าผลลัพธ์การกรองจากแต่ละส่วน ซึ่งได้ค่านวณใน $D_{i,l_0-\lambda}$ ด้วยสมการที่ 4.4

ต่อมาสามารถเลือกตำแหน่งที่ดีที่สุดสำหรับแต่ละส่วนได้อย่างอิสระ เพราะไม่มีปฏิสัมพันธ์ระหว่างส่วนที่มีผลต่อการคิดคะแนนรวม นอกจากการคำนวณ $D_{i,l_0-\lambda}$ แล้วนั้น อัลกอริทึมจากงานวิจัย [11] ยังสามารถคำนวณการกระจัดที่เหมาะสมที่สุดสำหรับแต่ละส่วน เพื่อหาตำแหน่งยึดของแต่ละส่วน จากสมการที่ 4.6

$$P_{i,l}(x, y) = \arg \max_{dx, dy} (R_{i,l}(x + dx, y + dy) - d_i \cdot \phi_d(dx, dy)) \quad (4.6)$$

หลังจากได้ทำการหาตำแหน่ง (x_0, y_0, l_0) ของ root ที่มีค่าคะแนนสูง สามารถหาตำแหน่งยึดของแต่ละส่วนที่สอดคล้องกับตำแหน่งของ root จากการหาการกระจัดที่เหมาะสมที่สุดจาก $P_{i,l_0-\lambda}(2(x_0, y_0) + v_i)$

4.2.2 Latent SVM

พิจารณาตัวแบ่งกลุ่มที่ให้คะแนนตัวอย่าง x ด้วยฟังก์ชัน :

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \quad (4.7)$$

$$\beta = (F'_0, \dots, F'_n, d_1, \dots, d_n, b) \quad (4.8)$$

กำหนดให้ β คือตัวแปรเวกเตอร์ของโมเดลซึ่งประกอบไปด้วย โมเดลตัวกรอง root (F'_0) โมเดลตัวกรองส่วน (F'_i) เวกเตอร์ (d_i) และค่า bias (b) และ z คือค่าแฝง (latent value) กำหนดให้เซต $Z(x)$ คือกลุ่มของค่าแฝงที่เป็นไปได้สำหรับกลุ่มตัวอย่าง x การแบ่งกลุ่มสำหรับ x สามารถพิจารณาได้จากการขีดแบ่งค่าคะแนน

โดยทั่วไปของ SVM จะสามารถฝึกสอน β จาก การกำหนดคำตอบของตัวอย่าง $D = ((x_1, y_1), \dots, (x_n, y_n))$ โดย $y_i \in \{-1, 1\}$ ด้วยการพิจารณาหาค่าสมการเป้าหมาย (objective function) ที่มีค่าต่ำสุด

$$L_D(\beta) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i f_\beta(x_i)) \quad (4.9)$$

โดย $\max(0, 1 - y_i f_\beta(x_i))$ เป็นฟังก์ชัน hinge loss และค่าคงที่ C ควบคุมน้ำหนักที่สัมพันธ์กับ Regularization Term

ถ้ามีค่าแฝงที่เป็นไปได้ค่าเดียวสำหรับแต่ละตัวอย่าง ($|Z(x_i)| = 1$) แล้ว f_β เป็น linear ใน β และ linear SVM เป็นกรณีพิเศษของ latent SVM

4.3 การหาแตร็คเล็ทโดยวิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด [2]

จากขั้นตอนที่ 4.2 การตรวจจับวัตถุด้วยโมเดลฝึกสอนแบบแยกส่วน สำหรับในขั้นตอนนี้จะเป็นการนำผลลัพธ์การตรวจจับคนมาทำการเชื่อมโยงข้อมูลเพื่อสร้างแตร็คเล็ท (แตร็คเล็ท หมายถึงเส้นทางการเดินในช่วงสั้นๆ) โดยเริ่มจากการแบ่งข้อมูลวิดีโอ ออกเป็นช่วงย่อย (segment) ตามจำนวนที่กำหนด และทำการหาแตร็คเล็ทในแต่ละช่วงย่อย โดยใช้วิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุดมาใช้ในการหาแตร็คเล็ท

เริ่มต้นจากการแบ่งวิดีโอออกเป็นจำนวน s ช่วงย่อย (segment) แต่ละช่วงย่อยจำนวน f เฟรม งานวิจัย[2] ได้เสนอวิธีการเชื่อมโยงข้อมูลสำหรับการหาแตร็คเล็ทซึ่งเป็นการหาจากการเคลื่อนไหว (Motion Cost) และรูปร่างพรรณสัมพันธ์ (Appearance Cost) ที่สอดคล้องกันจากทั้งช่วงย่อยของวิดีโอ โดยข้อมูลนำเข้าสำหรับการเชื่อมโยงข้อมูลเพื่อหาแตร็คเล็ทอยู่ในรูปแบบกราฟ $G = (V, E, w)$ โดยกำหนดให้ V เป็นเซตของโหนด E เป็นเซตของขอบ และ w เป็นน้ำหนักของขอบ โดย V แบ่งออกเป็นจำนวน f กลุ่ม ซึ่งแต่ละกลุ่มหมายถึงหนึ่งเฟรม และโหนดที่อยู่ข้างในหมายถึงการตรวจจับบุคคลในเฟรมนั้นๆ ให้ C_i คือ เฟรมที่ i (\equiv กลุ่ม) โดย $i \in \mathbb{Z} : 1 \leq i \leq f$ และ ให้ v_m^i หมายถึง การตรวจจับบุคคลที่ m (\equiv โหนด) ในเฟรมที่ i ดังนั้น $C_i = \{v_1^i, v_2^i, v_3^i, \dots\}$ กำหนดให้ขอบของกราฟคือ $E = \{(v_m^i, v_n^j) \mid i \neq j\}$ ซึ่งมีความหมายว่าทุกโหนดใน G จะเชื่อมกันได้ครบเท่าที่แต่ละโหนดไม่ได้อยู่ในกลุ่มเดียวกัน v_m^i เป็นตัวแทนของคุณลักษณะตาแห่ง x_m^i เป็นพิกัด 2 มิติของตาแห่งตรงกลางของบุคคลที่ตรวจจับได้ และคุณลักษณะ Appearance $\phi_{m,l}^i$ เป็นค่าฮิสโตแกรมสีของการแบ่งส่วนร่างตาแห่งที่ l ของการตรวจจับ v_m^i คำนวณน้ำหนักของขอบระหว่างสองโหนดจะเป็น $w : E \rightarrow \mathbb{R}^+$ ซึ่งคือค่าความคล้ายคลึงระหว่างการตรวจจับบุคคลสองบุคคล โดยสามารถคำนวณได้จากสมการที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$w(v_m^i, v_n^j) = \sum_{l=1}^8 k(\phi_m^i, \phi_n^j) \quad (4.10)$$

โดย k หมายถึง การอินเตอร์เซกชันของฮิสโตแกรม (histogram intersection kernel)

ในการหาแตร็คเล็ทของบุคคลคนเดียวกันในแต่ละส่วนย่อยของวิดีโอจำเป็นต้องมีการระบุตัวบุคคลที่ตรวจจับได้ในแต่ละเฟรม ดังนั้น คาตอบที่เป็นไปได้ของปัญหานี้จะถูกนำเสนอในรูปแบบของกราฟย่อยของกราฟ G ซึ่งเป็นกราฟที่มีหนึ่งโหนด (\equiv การตรวจจับ) ที่ถูกเลือกจากแต่ละกลุ่ม (\equiv เฟรม) กำหนดให้กราฟย่อยนี้เป็นคาตอบที่เป็นไปได้อยู่ในรูป $G_s = (V_s, E_s, w_s)$ ดังนั้น G_s จะเก็บเซตของโหนดที่นำเข้ามาจาก $V_s = \{v_a^1, v_b^2, v_c^3, \dots\}$ ซึ่งหมายถึงโหนดที่ a จากกลุ่มที่ 1 โหนดที่ b จากกลุ่มที่ 2 และโหนดอื่นที่ถูกเลือกจะอยู่ใน V_s

เป้าหมายของวิธีการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด คือการหาคาตอบที่เป็นไปได้โดยเลือกค่าน้อยที่สุด โดยค่าดังกล่าวคือค่าผลรวมของน้ำหนักทั้งหมดในกราฟย่อยของคาตอบที่เป็นไปได้ ในการที่จะรวมคุณลักษณะ Motion เข้ากับคุณลักษณะ Appearance เพื่อใช้ในการเชื่อมโยงข้อมูล สามารถคำนวณได้จากสมการ 4.11

$$\hat{V}_s = \underset{V_s}{\operatorname{argmin}} (\gamma_{\text{appearance}}(V_s) + \alpha \cdot \gamma_{\text{motion}}(V_s)) \quad (4.11)$$

โดย \hat{V}_s เป็นคาตอบที่เหมาะสมที่สุดของการเชื่อมโยงสำหรับหนึ่งแตร็คเล็ท และ α คือค่าคงที่ของการผสมเพื่อทำให้การรวมค่า Appearance และค่า Motion มีความสมดุลกัน การหา \hat{V}_s จากในสมการที่ 4.11 จะให้ผลลัพธ์เป็นแตร็คเล็ทของคนหนึ่งคนใน ส่วนย่อยของวิดีโอ ดังนั้นในการที่จะหาแตร็คเล็ทของคนใน ส่วนย่อยของวิดีโอ จะต้องคำนวณตามสมการที่ 4.11 ซ้ำอีกหลายครั้ง โดยในครั้งแรกที่สมการที่ 4.11 คำนวณเสร็จ อัลกอริทึมจะหาแตร็คเล็ทที่มีค่ารวมน้อยที่สุด ซึ่งนั่นคือรูปแบบคุณลักษณะ Appearance และ Motion มีความเสถียรที่สุด จากนั้น แตร็คเล็ทที่ดีที่สุดที่ถูกเลือกแล้วใน V_s จะถูกคัดออกมาจาก G และทำการหาแตร็คเล็ทซ้ำสำหรับของคนถัดไป และจะทำซ้ำจนกว่าใน G จะเหลือจำนวน 0 โหนดหรือเหลือแต่โหนดที่ใช้ไม่ได้

กำหนดให้ค่า Appearance ของคาตอบที่เป็นไปได้ V_s สามารถคำนวณได้จากสมการ 4.12

$$\gamma_{\text{appearance}}(V_s) = \frac{1}{2} \left(\sum_{i=1}^f \sum_{j=1, j \neq i}^f w(V_s(i), V_s(j)) \right) \quad (4.12)$$

การหาคาตอบที่เป็นไปได้โดยใช้ค่า Appearance ที่น้อยที่สุด นั่นคือ $\underset{V_s}{\operatorname{argmin}} (\gamma_{\text{appearance}}(V_s))$ ทำให้สามารถตรวจพบแตร็คเล็ทของบุคคลมีค่าฮิสโตแกรมที่มีความเสถียรในแต่ละส่วนย่อยของวิดีโอได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการคำนวณค่า Motion เพื่อไปใช้ในการคำนวณในสมการที่ 4.11 ต้องคำนวณค่าโดยอยู่บนพื้นฐานของการเคลื่อนไหวเพื่อใช้หาค่าตอบที่เป็นไปได้ V_s โดยเวกเตอร์ความเร็วเชิงตำแหน่งของคำตอบที่เป็นไปได้ V_s กำหนดให้อยู่ในรูป $\dot{X}_s(i) = X(i+1) - X(i)$ โดย $1 \leq i \leq (f-1)$ วิธีการหนึ่งที่สามารถใช้คำนวณค่า Motion นั้นคือการคำนวณค่าความเบี่ยงเบนจากรูปแบบที่สันนิษฐาน เช่น ใช้ความเร็วคงที่ ซึ่งสามารถทำได้โดยใช้เวกเตอร์ความเร็วของแต่ละคนมา ทารท นายต นแห่งของการตรวจจับถัดไป หลังจากนั้นก็ทำการหาค่าความผิดพลาดระหว่างตำแหน่งที่ได้จากการทำนายกับตำแหน่งในชุดคำตอบที่สามารถเป็นไปได้ ในการคำนวณนั้นได้ตั้งข้อสันนิษฐานว่ารูปแบบการเดินนั้นมีความเร็วที่คงที่ตลอดในส่วนย่อยของวิดีโอและจะสามารถคำนวณคุณลักษณะ Motion ได้จากสมการ 4.13 กำหนดให้ค่าคุณลักษณะ Motion ของค ตอบที่เป็นไปได้ V_s คือ

$$Y_{motion}(V_s) = \sum_{i=1}^s \sum_{j=1}^{s-1} \frac{\text{deviation}}{\text{prediction}} |X_s(i) - [X_s(j) + \dot{X}_s(j) \cdot (i-j)]| \quad (4.13)$$

จัดการปัญหาการโดนบดบังด้วยโหนดสมมุติ (Hypothetical Nodes) ในบางกรณี เฟรมที่พิจารณาอาจไม่มีผลการตรวจจับบุคคลเนื่องจากคนๆนั้นโดนบดบัง การตรวจจับผิดพลาด และสาเหตุอื่นๆ ซึ่งในประเด็นนี้สามารถรับมือได้ด้วยการเพิ่มโหนดสมมุติไปสู่แต่ละกลุ่ม ดังนั้นถ้าเฟรมใดเฟรมหนึ่งไม่มีการตรวจจับที่เหมาะสม โหนดสมมุติจะเป็นผู้ถูกเลือกแทน

จำเป็นต้องมีการกำหนดค่าคุณลักษณะ Appearance และค่าคุณลักษณะ Motion ทั้งสองค่าให้กับโหนดสมมุติ ทุกโหนดใน G จากนั้นการแก้ปัญหาเพื่อหาค่าที่เหมาะสมจะเป็นกระบวนการซ้ำตามสมการที่ 4.11 ในแต่ละรอบทั้งสองค่านี้จะถูกค นวณใหม่ให้กับโหนดสมมุติ ซึ่งในส่วนของการคำนวณและการปรับปรุงโหนดสมมุติจะขออธิบายในส่วนถัดไป

V_s ถูกคาดหวังให้เป็นคำตอบที่เป็นไปได้การติดตามของบุคคลหนึ่งคน (โหนดที่อยู่ในเกณฑ์ที่ถูกต้อง (inlier)) และการตรวจจับอื่นๆ (โหนดที่ผิดพลาด (outlier)) ในเฟรมต่างๆ ที่ไม่สามารถตรวจจับของคนๆนั้นได้ เนื่องจากส่วนย่อยของวิดีโอเป็นช่วงเวลาสั้นๆ จึงสามารถตั้งสมมุติฐานว่าการเคลื่อนไหวของคนจะเคลื่อนไหวด้วยความเร็วคงที่ตลอดทั้งส่วนย่อยของวิดีโอและจะใช้สมมุติฐานนี้ในการแยกแยะว่าเป็นค่าที่อยู่ในเกณฑ์หรือค่าที่ผิดปกติใน V_s อีกด้วย ตามที่การเคลื่อนไหวด้วยความเร็วคงที่นั้นอยู่ในรูปฟังก์ชันสองมิติ ตำแหน่งของการตรวจจับสามารถจำลองขึ้นโดยใช้ $X_s(i) = a_1 i + a_0$ โดยที่ a_1 และ a_0 เป็นเวกเตอร์คงที่ในระบบสองมิติ ดังนั้นสามารถระบุค่าที่อยู่ในเกณฑ์และค่าที่ผิดปกติด้วยก าหนดรูปแบบควมเร็วคงที่จาก a_1 และ a_0 โดยใช้

$$V_s^{inliers} = \{V_s(i): |a_1 i + a_0 - X_s(i)| < \delta\} \quad (4.14)$$

โดยที่ δ คือค่าเงื่อนไขการเบี่ยงเบนที่เหมาะสม โดยตัวแปรที่ดีที่สุดของโมเดลความเร็วคงที่ของแทร็คเล็ตต้องเป็นรูปแบบซึ่งจะหาให้ได้ จำนวนแทร็คเล็ตที่อยู่ในเกณฑ์มากที่สุด

$$\hat{a}_1, \hat{a}_0 = \underset{a_1, a_0}{\operatorname{argmax}} (\# \{V_s^{\text{inliers}}\}) \quad (4.15)$$

ให้ $\#$ หมายถึง จำนวนสมาชิกของเซต เนื่องจาก V_s ประกอบด้วยค่าที่อยู่ในเกณฑ์และค่าที่ผิดปกติ งานวิจัย[2] ใช้เทคนิค RANSAC เพื่อจะได้ค่า a_1 และ a_0 ตามในสมการที่ 4.15 และไปหาเกณฑ์ความผิดพลาดต่อในสมการที่ 4.14

โหนดที่อยู่ในเกณฑ์ใน V_s คือโหนดที่ผ่านการคัดเลือกจากสมการที่ 4.14 โดยใช้ \hat{a}_1 และ \hat{a}_0 โดยตำแหน่งพิกัดของโหนดสมมุติเป็นการคำนวณจากสมการ $x_H^i = \hat{a}_1 i + \hat{a}_0$ โดย x_H^i หมายถึง พิกัดของโหนดสมมุติของกลุ่มที่ i สำหรับค่าคุณลักษณะ Appearance ของโหนดสมมุติจะเป็นค่าเฉลี่ยของคุณลักษณะ Appearance ของ inlier (เป็นค่าเฉลี่ยของฮิสโตแกรมสี) แต่อย่างไรก็ตาม ค่าปรับคงที่จะถูกเพิ่มกับค่านำหนักของเส้นขอบที่เชื่อมโยงโหนดสมมุติใน G แต่จะพยายามหลีกเลี่ยงการเลือกโหนดสมมุติถ้าหากในเฟรมนั้นมีการตรวจจับที่เหมาะสมอยู่

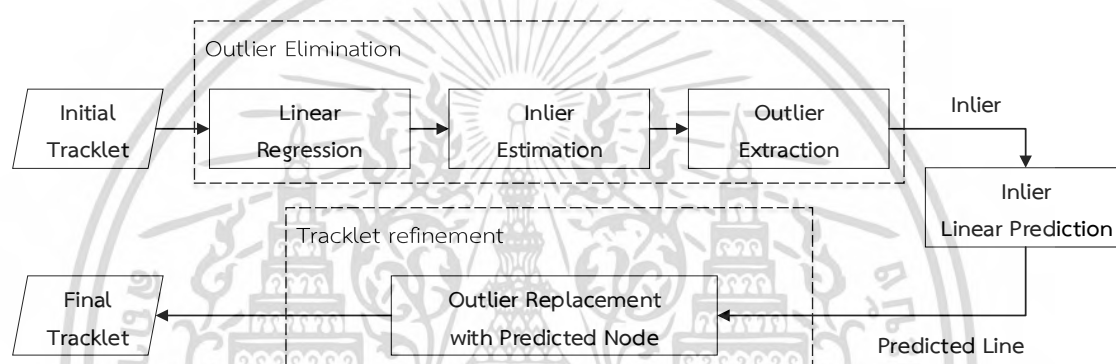
4.4 การปรับปรุงแทร็คเล็ตด้วยการประมาณตำแหน่ง

ผลลัพธ์ของระบบการติดตามบุคคล (Human Tracking) นั้น ย่อมมีทั้งผลลัพธ์ที่เชื่อมโยงถูกต้องเป็นคนเดียวกันและมีทั้งผลลัพธ์ที่เลือกการตรวจจับที่ไม่ใช่คนเดียวกันมาเป็นคำตอบ สาเหตุเนื่องมาจากการตรวจจับบุคคลที่ผิดพลาด หรือการที่คนๆนั้นโดนบดบังอยู่ทำให้ไม่มีผลการตรวจจับบุคคลที่ต้องการ ส่งผลให้ในกระบวนการเชื่อมโยงข้อมูลหยาบเลือก ทอปปที่ผิดพลาดมาเป็นผลลัพธ์ของแทร็คเล็ต งานวิจัย [2] แก้ปัญหาการเชื่อมโยงข้อมูลที่ผิดพลาดด้วยการสร้างโหนดสมมุติ เพิ่มไปในแต่ละกลุ่มเพื่อให้โหนดสมมุติเป็น 1 ในตัวเลือก ทำให้สามารถเชื่อมโยงหาแทร็คเล็ตได้มีประสิทธิภาพมากขึ้น โดยงานวิจัย [2] ได้ตั้งสมมุติฐานว่าการเคลื่อนไหวของคนจะเคลื่อนไหวด้วยความเร็วคงที่และเป็นเส้นตรงตลอดทั้งส่วนย่อยของวิดีโอ โดยการพิจารณาเชื่อมโยงข้อมูลเพื่อหาแทร็คเล็ตนั้นจะเกิดภายในแต่ละส่วนย่อยของวิดีโอ แล้วจึงค่อยหาแทร็คเล็ตไปรวมเพื่อหาเส้นทางการเดิน ดังนั้นจากสมมุติฐานนี้ย่อมหมายถึงว่า แต่ละแทร็คเล็ตจะมีลักษณะที่คนเดินเป็นเส้นตรง และเดินด้วยความเร็วที่คงที่ จากงานวิจัย [2] ในส่วนขั้นตอนการสร้างโหนดสมมุติ ได้เลือกวิธีการหาโหนดที่อยู่ในเกณฑ์ (inlier) ด้วยวิธีการ RANSAC โดยจะพิจารณาหา inlier ที่แกน sequence-x และแกน sequence-y แยกจากกันอิสระ แล้วจึงค่อยหา inlier ที่เหมือนกันจากทั้งสองแกน ไปทำการพิจารณาสร้างโหนดสมมุติต่อไป จากรูปที่ 4.6 (ค) พบว่าการแบ่ง inlier และโหนดที่ผิดพลาด (outlier) แบบพิจารณาแยกแกนแบบที่ [2] นาเสนอ ยังไม่สามารถสกัด inlier ได้อย่างถูกต้อง ยังคงมี inlier และ outlier ผสมปะปนกันอยู่ในแนววิถีเส้นตรง ซึ่งจะมีผลให้วิถีเส้นตรงที่ประมาณมีความผิดพลาดได้ ทำให้ได้ผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลัพท์แทร็คเก็ตที่ไม่ถูกต้องในบางเฟรม งานวิจัยนี้จึงนำเสนอการปรับปรุงแทร็คเก็ตให้มีความถูกต้องมากยิ่งขึ้น ด้วยวิธีการสร้างเส้นตรงตัวแทนของแทร็คเก็ต (Predicted Line) และปรับปรุงประสิทธิภาพของแทร็คเก็ตด้วยโหนดจากเส้นตรงตัวแทนของแทร็คเก็ต (Predicted Node) เพื่อให้เกิดความแม่นยำ ขาของแทร็คเก็ตมากยิ่งขึ้น

ขั้นตอนในส่วนของการปรับปรุงแทร็คเก็ตของงานวิจัยที่นำเสนอนี้ เป็นการนำเสนอ วิธีการปรับปรุงความแม่นยำของแทร็คเก็ตที่เชื่อมโยงข้อมูลของคนที่เดินด้วยความเร็วคงที่และเป็นเส้นตรง ให้มีความแม่นยำมากยิ่งขึ้น ด้วยการสร้างเส้นตรงตัวแทนของแทร็คเก็ตขึ้นมาและนำเส้นตัวแทนมาปรับปรุงแทร็คเก็ตให้มีความแม่นยำ มากยิ่งขึ้น



รูปที่ 4.5 ขั้นตอนการปรับปรุงแทร็คเก็ต

จากรูปที่ 4.5 งานวิจัยที่นำเสนอ จะแบ่งขั้นตอนออกเป็น 2 ขั้นตอนได้แก่ การกำจัดโหนดผิดพลาด (Outlier Elimination) และการปรับปรุงแทร็คเก็ต (Tracklet refinement) สำหรับผลการติดตามบุคคลที่มีความผิดพลาดนั้น ถ้าหากนำแทร็คเก็ตมาพล็อตบนแกน x-y แล้ว จะพบว่าจุดที่ผิดพลาดจะมีตำแหน่งที่โดดออกไปจากวิถีเส้นตรงของแนวแทร็คเก็ต ในขั้นตอนของการกำจัดโหนดผิดพลาด จะเป็นการนำเสนอการกำจัดผลการเชื่อมโยงข้อมูลที่ผิดพลาดออก และนำเฉพาะ inlier ที่ได้ไปหาการประมาณค่าเส้นตรงตัวแทนของแทร็คเก็ต เพื่อคาดคะเนตำแหน่งที่ถูกต้องในแต่ละเฟรมขึ้นมา แล้วจึงนำโหนดบางโหนดจากเส้นตรงตัวแทนแทร็คเก็ตไปทดแทนในส่วนที่เป็น outlier ของแทร็คเก็ตดั้งเดิม เพื่อให้ได้แทร็คเก็ตที่มีความถูกต้องแม่นยำ มากยิ่งขึ้น

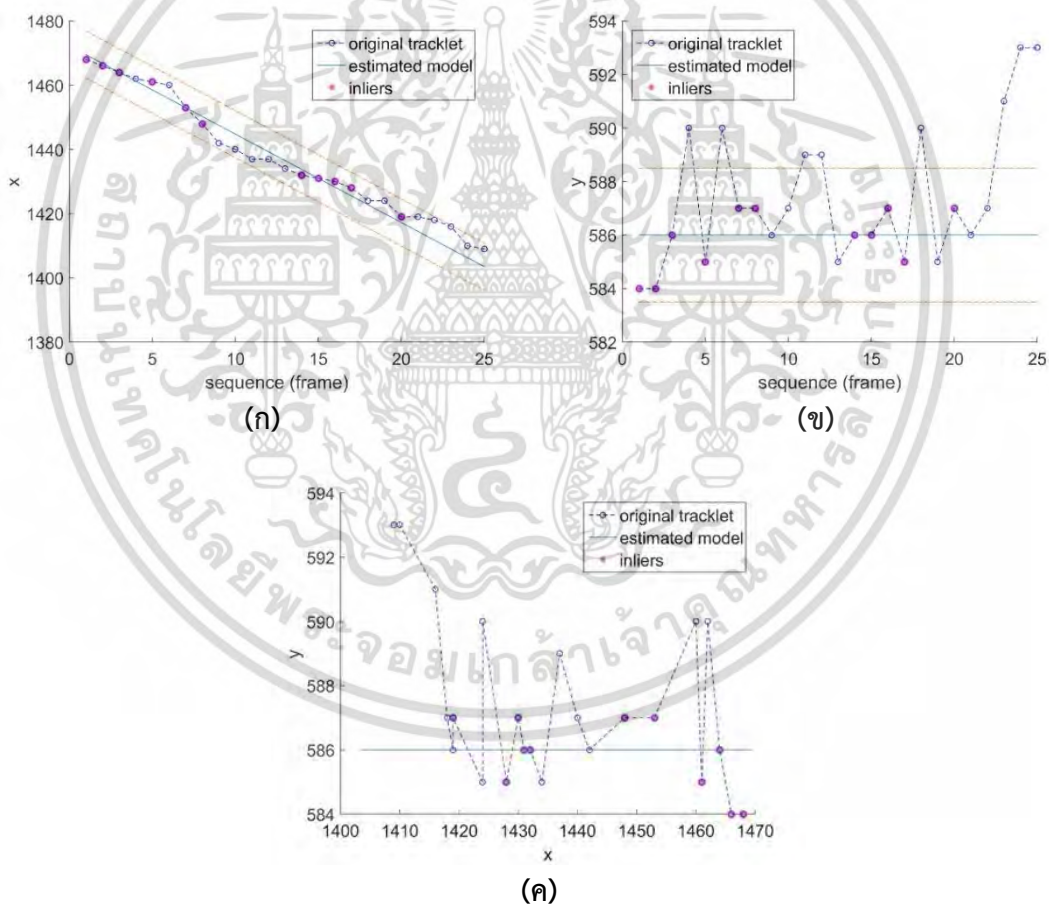
4.4.1 การกำจัดโหนดผิดพลาด (Outlier Elimination)

ในงานวิจัยนี้ ตำแหน่งของการตรวจจับในขั้นตอนของการติดตามคน จะถูกใช้ในที่ที่ ตำแหน่งจุดกึ่งกลางของสี่เหลี่ยมล้อมรอบ ซึ่งถ้าหากผลลัพธ์การตรวจจับคนมีความผิดพลาด เช่น ตีกรอบใหญ่เกินตัวคน เล็กกว่าตัวคน หรือมีตำแหน่งที่ไม่ตรงกับตัวคน จะทำให้จุดกึ่งกลางที่ได้ไม่ตรงกับจุด

กึ่งกลางของตัวคนจริงๆ ซึ่งความผิดพลาดของตำแหน่งสามารถส่งผลกระทบต่อประสิทธิภาพของผลลัพธ์การติดตามบุคคลได้

แทร็คเก็ตที่มีการเชื่อมโยงข้อมูลที่ผิดพลาดนั้น หากนำลำดับจุดกึ่งกลางของหน้าต่างการตรวจจับทั้งหมดมาพิจารณาบนแกน $x-y$ จะพบว่า ตำแหน่งของการตรวจจับที่ผิดพลาดจะมีระยะที่อยู่นอกเกณฑ์ของวิถีเส้นตรงจากตำแหน่งการตรวจจับอื่นๆ ภายในแทร็คเก็ต ซึ่งถ้าหากนำแทร็คเก็ตที่ยังมีการตรวจจับที่ผิดพลาดปะปนอยู่ ไปทำการประมาณเส้นตรงตัวแทนแทร็คเก็ตโดยตรงเลย อาจจะทำให้เส้นตรงตัวแทนแทร็คเก็ตที่ได้มีความผิดพลาดเพิ่มขึ้นไป

การปรับปรุงประสิทธิภาพของการติดตามวัตถุให้ดียิ่งขึ้น มีความจำเป็นที่โหนดผิดพลาด (outlier) จากแทร็คเก็ตเดิมจะต้องถูกกำจัดและควรจะสกัดเหลือแค่เพียงโหนด inlier ก่อนนำไปประมาณหาเส้นตรงตัวแทนแทร็คเก็ต เพื่อใช้ในการปรับปรุงแทร็คเก็ต



รูปที่ 4.6 การหา inlier จาก tracklet บนแกนต่างๆ

(ก) การหา inlier จาก tracklet แยกประมาณในแกน sequence (frame) - X

(ข) การหา inlier จาก tracklet แยกประมาณในแกน sequence (frame) - Y

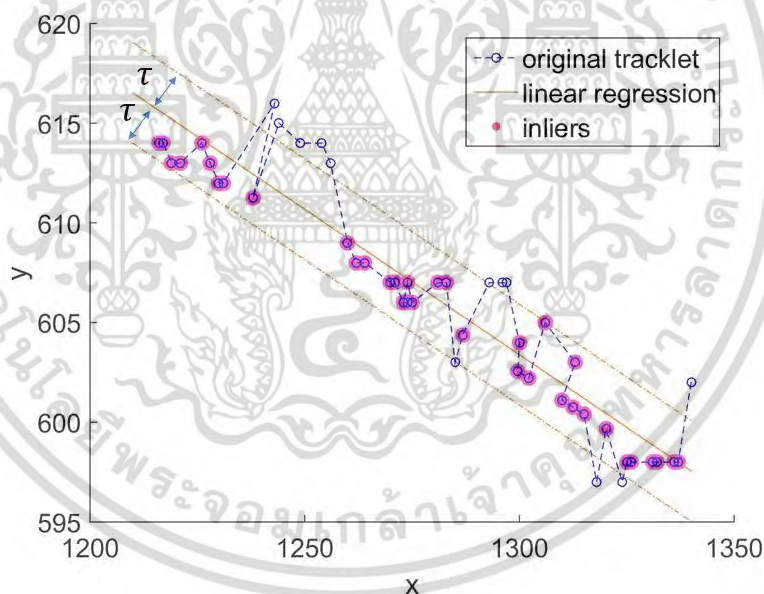
(ค) inlier บน tracklet ที่นำผลจากการแยกประมาณในแกน X (ก) และ

การแยกประมาณในแกน Y (ข) มารวมกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากงานวิจัย [2] ในส่วนขั้นตอนการสร้างโหนดสมมุติ ได้เลือกวิธีการหา inlier ด้วยวิธีการ RANSAC โดยจะพิจารณาหา inlier ที่แกน sequence-x และแกน sequence-y แยกจากกันอิสระ แล้วจึงค่อยหา inlier ที่เหมือนกันจากทั้งสองแกน ไปทำ การพิจารณาสร้างโหนดสมมุติต่อไป จากรูปที่ 4.6 (ค) พบว่าการแบ่ง inlier และ outlier แบบพิจารณาแยกแกนแบบนี้ [2] น่าเสนอนั้น ยังไม่สามารถสกัด inlier ได้อย่างถูกต้อง ยังคงมี inlier และ outlier ผสมปะปนกันอยู่ในแนววิถีเส้นตรง ซึ่งจะมีผลให้วิถีเส้นตรงที่ประมาณมีความผิดพลาดได้

งานวิจัยที่น่าเสนอ ในขั้นตอนของการกำจัด outlier จะพิจารณาที่แกน x-y พร้อมกันทีเดียว โดยในขั้นตอนนี้เลือกใช้วิธีการ linear regression แบบ least squares ในการประมาณแนวเส้นของแทร็คเคลื่อนออกมา หลังจากนั้นจึงค่อยหา inlier และ outlier โดยเกณฑ์ในการแบ่งคือ ระยะห่างระหว่างจุดกึ่งกลางของแต่ละผลลัพธ์การตรวจจับกับแนวเส้นของแทร็คเคลื่อนที่ที่ประมาณ โดยกำหนดให้ระยะห่างระหว่างจุดกับแนวเส้นตรงของแทร็คเคลื่อนที่ไม่เกินค่า τ จะถือว่าเป็น inlier ส่วนจุดที่มีระยะห่างมากกว่าค่า τ จะถือว่าเป็น outlier หลังจากกำจัด outlier แล้วจึงค่อยหา inlier ที่ได้ไปพิจารณาหาเส้นตรงตัวแทนของแทร็คเคลื่อนที่ในขั้นตอนต่อไป



รูปที่ 4.7 การหา inlier บน tracklet ด้วยวิธีการ linear regression แบบ least squares ในแกน x-y

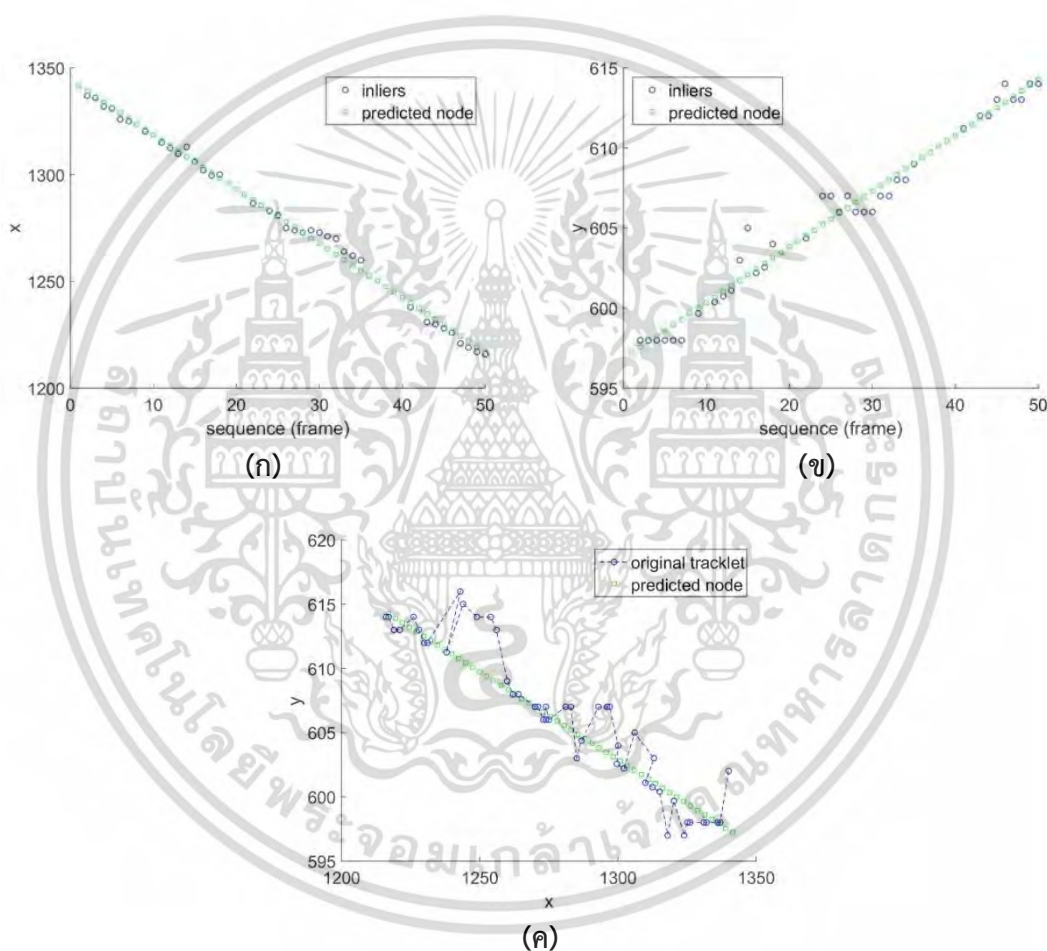
เมื่อได้ inlier มาแล้ว inlier เหล่านั้นเป็นค ทอบที่มีแนวโน้มที่จะเป็นคนอื่นๆเดียวกันภายในแทร็คเคลื่อนที่ จากงานวิจัยที่น่าเสนอ ได้ตั้งสมมุติฐานว่าแทร็คเคลื่อนที่จะมีลักษณะที่เป็นเส้นตรงและคนเดินด้วยความเร็วคงที่ ดังนั้นการจะสร้างเส้นตรงตัวแทนแทร็คเคลื่อนที่ขึ้นมา จะต้องเป็นการสร้างโมเดลให้อยู่ในรูปสมการเส้นตรง โดยใช้วิธีการ linear regression แบบ least squares ในการสร้างรูปแบบจาสองการเคลื่อนที่ด้วยความเร็วคงที่ของคนเดินในแทร็คเคลื่อนที่นั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P_X^i = \hat{w}_{x1}i + \hat{w}_{x0} \quad (4.16)$$

$$P_Y^i = \hat{w}_{y1}i + \hat{w}_{y0} \quad (4.17)$$

จากสมการ 4.16 และ 4.17 กำหนดให้ P_X^i คือ ตำแหน่งของโหนดที่ได้จากการทำนายในแนวแกน X ในเฟรมที่ i , P_Y^i คือ ตำแหน่งของโหนดที่ได้จากการทำนายในแนวแกน Y ในเฟรมที่ i , \hat{w}_{x1} , \hat{w}_{x0} คือค่า regression coefficient สำหรับการทำนายในแนวแกน X , \hat{w}_{y1} , \hat{w}_{y0} คือค่า regression coefficient สำหรับการทำนายในแนวแกน Y

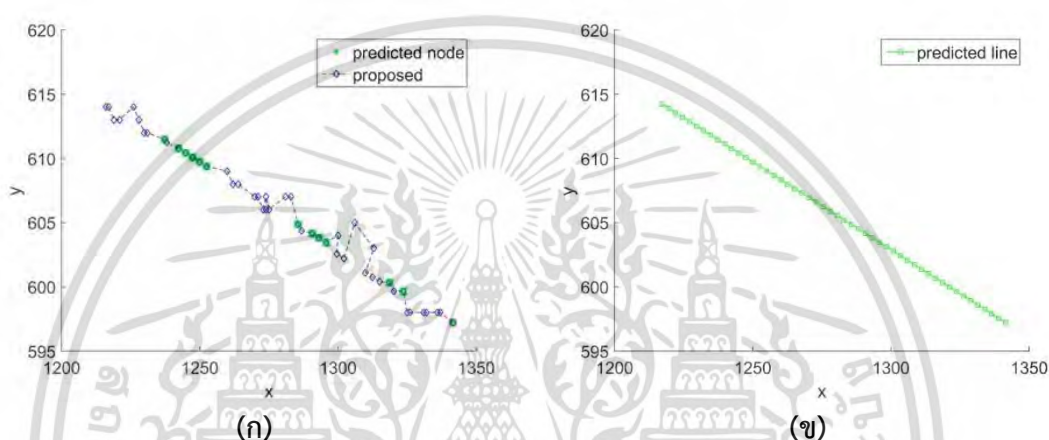


รูปที่ 4.8 เส้นตรงตัวแทนแตร็คเก็ตที่ทำนายจาก inlier บนแกนต่างๆ
 (ก) เส้นตรงตัวแทนแตร็คเก็ตที่ทำนายจาก inlier บนแกน sequence (frame) - X
 (ข) เส้นตรงตัวแทนแตร็คเก็ตที่ทำนายจาก inlier บนแกน sequence (frame) - Y
 (ค) เส้นตรงตัวแทนแตร็คเก็ตได้จาก (ก) และ (ข) มาแสดงบนแกน X - Y

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 การปรับปรุงแตร็คเล็ท

เพื่อให้ได้แตร็คเล็ทที่มีความแม่นยำมากยิ่งขึ้น งานวิจัยที่นำเสนอ จึงทำการปรับปรุงแตร็คเล็ทเดิม ด้วยการนำโหนดที่ได้จากเส้นตรงตัวแทนแตร็คเล็ท มาแทนที่โหนดของ outlier แตร็คเล็ทเดิมในเฟรมที่โหนดนั้นๆของแตร็คเล็ทเดิมเป็น outlier เพื่อให้ได้ผลของแตร็คเล็ทที่มีประสิทธิภาพมากยิ่งขึ้น โดยจากการวัดระยะห่างระหว่างจุด outlier กับ ground truth มีค่าเฉลี่ยระยะห่างเท่ากับ 7.99 ในขณะที่ระยะห่างระหว่างโหนดที่ได้จากเส้นตรงตัวแทนแตร็คเล็ท กับ ground truth มีค่าเฉลี่ยระยะห่างเท่ากับ 7.72 ซึ่งนับว่ามีระยะเฉลี่ยที่แม่นยำ ขึ้นถึง 3.37%



รูปที่ 4.9 แตร็คเล็ทที่ปรับปรุงแล้วและเส้นตรงตัวแทนแตร็คเล็ท

(ก) แตร็คเล็ทที่ปรับปรุงแล้ว

(ข) เส้นตรงตัวแทนแตร็คเล็ท

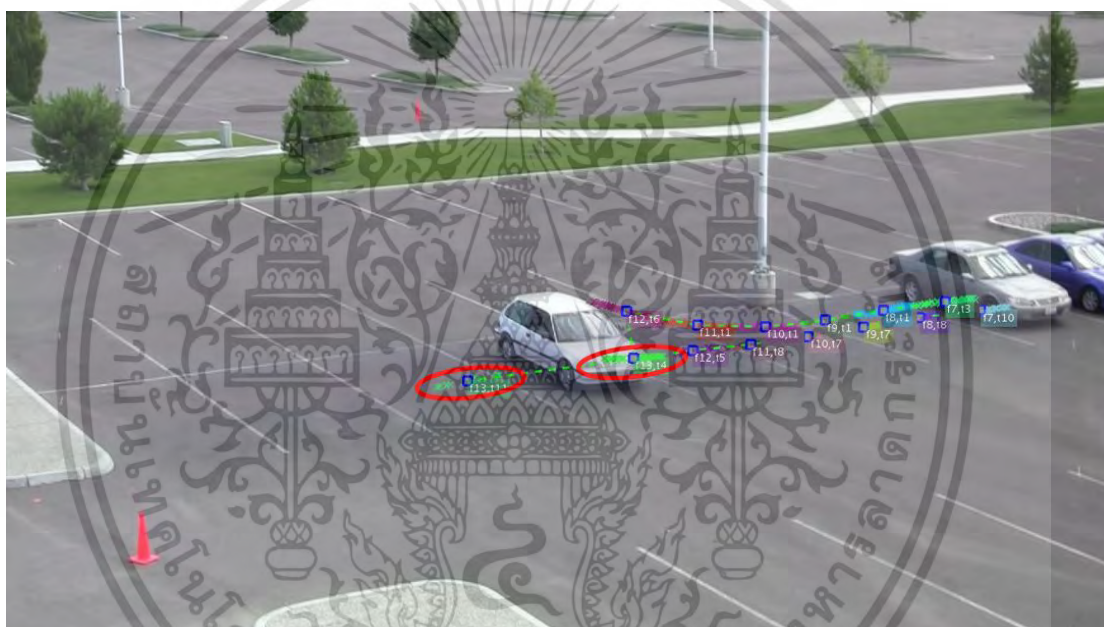
4.5 การรวมแตร็คเล็ทเป็นเส้นทางการเดินและปรับปรุงความถูกต้องด้วยเทคนิคการปรับเรียบข้อมูล

จากขั้นตอนที่ 4.3 การสร้างแตร็คเล็ท และนำแตร็คเล็ทที่ได้มาปรับปรุงให้มีคุณภาพที่ดีมากยิ่งขึ้นในขั้นตอนที่ 4.4 สำหรับในขั้นตอนนี้จะเป็นการรวมแตร็คเล็ทจากหลายส่วนย่อยของวิดีโอ ให้กลายเป็นเส้นทางการเดินตลอดทั้งวิดีโอ

จากงานวิจัย [2] ในขั้นตอนของการรวมแตร็คเล็ทเป็นเส้นทางการเดิน ได้ใช้คุณลักษณะรูปพรรณสัณฐาน (Appearance Cost) และค่าการเคลื่อนไหว (Motion Cost) ในการพิจารณา กำหนดให้กลุ่มและโหนดใน G' จะหมายถึงส่วนย่อยของวิดีโอและแตร็คเล็ทตามลำดับ สำหรับการเชื่อมโยงแตร็คเล็ทเป็นเส้นทางการเดิน จะต้องเลือกแตร็คเล็ทจากทุกส่วนย่อยของวิดีโอ มาส่วนละ 1 แตร็คเล็ท เพื่อมาเชื่อมโยงเป็นเส้นทางการเดิน โหนดแตร็คเล็ทที่ถูกเลือกมาเป็นเส้นทางการเดินจะถูกตัดทิ้งไปจากกลุ่ม ในแต่ละรอบของการหาเส้นทางการเดิน ถ้าหากกลุ่มใดไม่มีโหนดเหลือให้เลือกแล้ว กลุ่มนั้นก็จะเป็นที่ที่จะต้องนำมาพิจารณาอีก สำหรับใน 1 รอบของการหาหมายถึงการหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เส้นทางการเดินของคน 1 คน ในกรณีที่คนเดินถูกบดบัง หรือคนเดินออกนอกกรอบของวิดีโอไปแล้ว ส่งผลให้ในกลุ่มส่วนย่อยของวิดีโอไม่มีโหนดแทร็คเล็ทของคนดังกล่าว จากเงื่อนไขที่ต้องเลือกแทร็คเล็ทจากแต่ละกลุ่ม มากกลุ่มละ 1 โหนด เพื่อมาเป็นค ทอบของเส้นทางการเดิน ทำให้การเลือกคำตอบไปเลือกแทร็คเล็ทของบุคคลอื่นที่มีค่าคุณลักษณะใกล้เคียงกันมาเป็นค าสบแทน ซึ่งกรณีนี้ส่งผลให้ในรอบถัดๆไปที่ต้องเชื่อมโยงข้อมูลหาเส้นทางการเดินของคนที่ถูกแย่งแทร็คเล็ทไปเป็นคำตอบในรอบก่อนหน้านั้น ไม่มีคำตอบที่ถูกต้องให้เลือก เลยไปเลือกแทร็คเล็ทของคนอื่นมาเป็นคำตอบอีก แล้วก็ส่งผลกระทบเกิดการนาแทร็คเล็ทของบุคคลคนอื่นมาเป็นคำตอบ ซึ่งส่งผลกระทบเป็นลูกโซ่ ต่อการรวมแทร็คเล็ทของบุคคลต่อเนื่องกันไปกระทบแบบนี้ต่อไปเรื่อยๆ ในรอบถัดๆไป เช่นตัวอย่างในรูปที่ 4.10



รูปที่ 4.10 ผลเส้นทางการเดินที่มีการติดตามสลับตัวบุคคล

ถึงแม้ในงานวิจัย [2] ได้เสนอการสร้างโหนดสมมุติมาใช้ทดแทนแทร็คเล็ทของคนที่ไม่โดนบดบังหรือออกนอกกรอบของวิดีโอ เพื่อป้องกันปัญหาการติดตามสลับคนที่ได้กล่าวมาก่อนหน้านี้ แต่พบว่าปัญหาที่เกิดขึ้นจากการใช้โหนดสมมุติคือการเลือกแทร็คเล็ทจะไปเลือกแทร็คเล็ทสมมุติเป็นคำตอบมากเกินไป ในขณะที่ยังคงมีแทร็คเล็ทของบุคคลคนนั้นเป็นตัวเลือกอยู่ก็ตาม สาเหตุนี้เนื่องมาจากการสร้างโหนดสมมุติ จะต้องจำลองคุณลักษณะ Appearance และจำลองตาแหน่งของโหนดสมมุติ โดยใช้คำตอบเส้นทางการเดินที่ดีที่สุดที่ได้จากโหนดจริงมาใช้ในการจำลอง ค่าคุณลักษณะแทร็คเล็ทภายในคำตอบของเส้นทางการเดินที่ดีที่สุดที่ได้จากโหนดจริงจะถูกนาไปทานาย เพื่อสร้างโหนดสมมุติข้างหน้าและโหนดสมมุติข้างหลัง สิ่งที่เกิดขึ้นคือ การเลือกโหนดสมมุติทำให้ได้ค่าคุณลักษณะรวมที่ดีกว่าการไปเลือกโหนดจริง ถึงแม้จะยังมีโหนดจริงของบุคคลนั้นเป็น

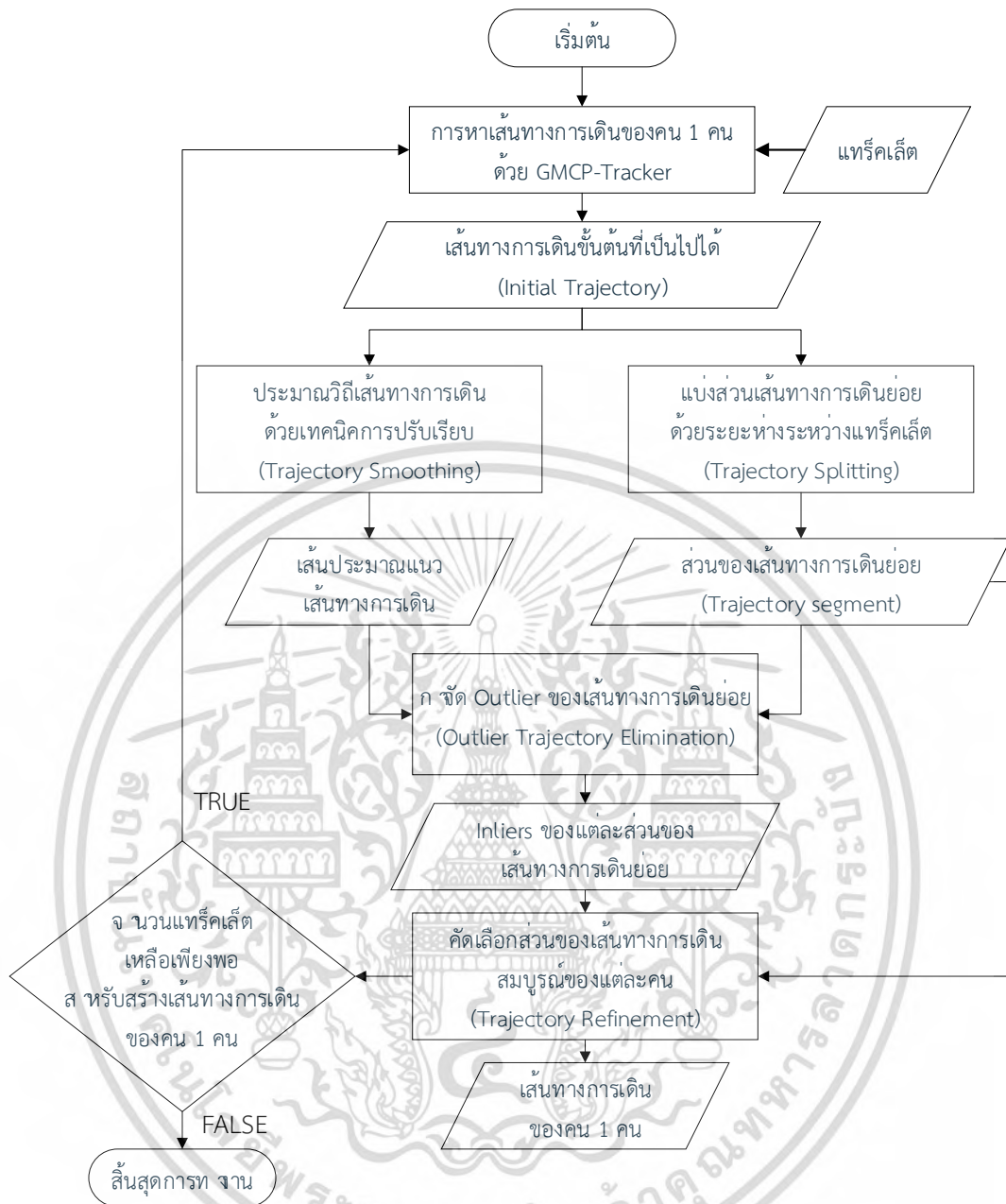
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวเลือกอยู่ก็ตาม ทำให้ผลคำตอบของเส้นทางการเดิน มีโหมดสมมุติถูกเลือกมาเป็นคำตอบมากเกินไป ความจำเป็น และโหมดจริงของบุคคลนั้นที่ไม่ถูกเลือก ยังคงอยู่เป็นตัวเลือกในกราฟ และถูกนำไปสร้างเป็นเส้นทางการเดินที่ผิดพลาดในรอบถัดไป

สำหรับงานวิจัยที่นำเสนอจะมุ่งเน้นไปที่การแก้ข้อจำกัดในการเลือกแตรัคเล็ทสลับตัวบุคคล เพื่อให้คำตอบของเส้นทางการเดินเป็นของคนเดียวกันตลอดทั้งคำตอบ สำหรับแนวคิดของงานวิจัยที่นำเสนอคือ การใช้เทคนิคการปรับเรียบข้อมูล (Smooth Data) มาทำการประมาณเส้นทางการเดิน เพื่อดูแนวโน้มและรูปแบบของวิถีเส้นทางการเดิน และนำเส้นที่ประมาณได้ไปแบ่งแยกเป็นส่วนของเส้นทางการเดินย่อย (Trajectory Segment) ช่วงใดที่มีลักษณะแปลกแยกไปจากวิถีการเดิน ส่วนที่เป็น outlier ที่มีลักษณะแปลกแยกห่างออกไปจากเส้นที่ประมาณได้ จะมีโอกาสเป็นแตรัคเล็ทที่สลับตัวบุคคลกัน ส่วนที่เป็น inlier ที่อยู่ในช่วงเส้นที่ประมาณได้ จะเป็นแตรัคเล็ทของบุคคลคนเดียวกันกับแตรัคเล็ทอื่นภายในคำตอบเส้นทางการเดิน งานวิจัยที่นำเสนอจะทำการคัดเลือกแตรัคเล็ทที่เป็น inlier มารวมเป็นคำตอบเส้นทางการเดิน และแตรัคเล็ทที่เป็น outlier จะถูกคืนไปเป็นตัวเลือกสำหรับพิจารณาสร้างเส้นทางการเดินในรอบถัดไป

จากรูปที่ 4.11 ขั้นตอนทั้งหมดในส่วนของการรวมแตรัคเล็ทเป็นเส้นทางการเดินและปรับปรุงความถูกต้องด้วยเทคนิคการปรับเรียบ ขั้นตอนแรกคือ ขั้นตอนการรวมแตรัคเล็ทเป็นเส้นทางการเดินขั้นต้น (Initial trajectory) โดยจะนำแตรัคเล็ทแต่ละส่วนย่อยของวิดีโอที่ได้จากขั้นตอนที่ 4.4 มาทำการรวมกันเป็นเส้นทางการเดิน ซึ่งในแต่ละรอบของการเชื่อมต่อแตรัคเล็ทจะหาแค่เพียงเส้นทางการเดินขั้นต้นของทีละ 1 คน ขั้นตอนต่อมาเป็นขั้นตอนการแยกส่วนเส้นทางการเดิน (Trajectory Splitting) โดยนำเส้นทางการเดินขั้นต้นที่ประมาณค่ามาได้นั้น ไปตัดแยกเป็น Trajectory segment โดยจะพิจารณาจากระยะห่างระหว่างแตรัคเล็ทแต่ละแตรัคเล็ทภายในคำตอบเส้นทางการเดินที่เป็นไปได้ ถ้าหากระยะห่างระหว่างแตรัคเล็ทมีระยะที่มากกว่าที่กำหนด จะมีโอกาสสูงที่จะเป็นแตรัคเล็ทของบุคคลคนละคนกัน จะถูกตัดแยกออกมาเป็น Trajectory segment ในขณะที่เดียวกันจะหาขั้นตอนของการประมาณวิถีเส้นทางการเดินด้วยเทคนิคการปรับเรียบ (Trajectory Smoothing) เป็นการหาเพื่อสร้างเส้นประมาณแนวเส้นทางการเดินควบคู่กันไป ซึ่งเป็นเหมือนการดูภาพรวมของเส้นทางการเดินว่ามีลักษณะวิถีการเดินเป็นอย่างไร หลังจากนั้นขั้นตอนต่อมาเป็นขั้นตอนการปรับปรุงเส้นทางการเดิน (Trajectory refinement) โดยจะพิจารณาว่าแต่ละส่วนของเส้นทางการเดินย่อยที่แบ่งมาได้ มีตำแหน่งที่อยู่ในบริเวณใกล้เส้นประมาณแนวเส้นทางการเดินมากน้อยเพียงใด Trajectory segment ที่ผิดพลาดที่มีระยะห่างจากวิถีเส้นทางการเดินที่ประมาณได้จะถูกตัดทิ้ง (Outlier Trajectory Elimination) เพื่อจะได้นำไปพิจารณาคัดเลือก Trajectory segment ที่อยู่ในวิถีมาเป็นคำตอบ และคืนส่วนของเส้นทางการเดินที่อยู่นอกวิถี ให้นำไปเป็นตัวเลือกไว้ใช้พิจารณาสำหรับสร้างเส้นทางการเดินของคนในรอบถัดๆไป เพื่อลดการดึงเชื่อมต่อกับเส้นทางผิดพลาดซึ่งอาจส่งผลกระทบต่อเป็นลูกโซ่ ต่อการรวมแตรัคเล็ทของบุคคลต่อเนื่องกันไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 ขั้นตอนการรวมแทร็คเล็ตเป็นเส้นทางการเดิน และปรับปรุงความถูกต้องด้วยเทคนิค smooth data

4.5.1 การรวมแตร็คเล็ทเป็นเส้นทางการเดินโดยใช้การพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุด [2]

จากที่ได้ทำการสร้างและปรับปรุงแตร็คเล็ทของคนเดินทุกคนจากแต่ละส่วนย่อยของวิดีโอที่ได้อธิบายในหัวข้อที่ 4.3 และ 4.4 ในขั้นตอนนี้จะเป็นการนำแตร็คเล็ทที่ได้มาสร้างเส้นทางการเดินของบุคคลที่สมบูรณ์จนครบทั้งวิดีโอ โดยใช้ปัญหาการพิจารณาตามเกณฑ์ของคลิกกราฟที่สั้นที่สุดเหมือนกับหัวข้อที่ 4.3 โหนดของกราฟจะถูกจับกลุ่มในรูปที่เป็นกลุ่มแยกอิสระ เพื่อหาซัพเซตของโหนดที่ประกอบด้วยเพียงหนึ่งโหนดจากแต่ละกลุ่ม ในขั้นตอนนี้กำหนดให้กลุ่มและโหนดใน G' หมายถึงส่วนย่อยของวิดีโอและแตร็คเล็ทตามลำดับ (แทนที่กลุ่มและโหนดหมายถึงเฟรมและการตรวจจับบุคคลตามลำดับในหัวข้อที่ 4.3 การสร้างแตร็คเล็ท) คุณลักษณะ Appearance ของโหนดที่เป็นตัวแทนของแตร็คเล็ท จะเป็นค่าที่ได้จากการเฉลี่ยค่า Appearance ของการตรวจจับบุคคลในแตร็คเล็ทนั้นๆ (เป็นค่าเฉลี่ยของฮิสโตแกรมสี) และ x'_m กำหนดให้เป็นตำแหน่งตรงกลางของแตร็คเล็ท

การเชื่อมโยงข้อมูลในระดับการหาเส้นทางการเดินนั้นมีความแตกต่างจากระดับในการหาแตร็คเล็ทจากที่ได้ตั้งข้อสมมุติฐานว่าการเดินของคนจะมีความเร็วคงที่ในหนึ่งส่วนย่อยของวิดีโอ แต่ในการจำลองการเคลื่อนไหวของคนในระยะเวลาอนานนั้น ในการติดตามจะมีความยากและซับซ้อนมากยิ่งขึ้น วิธีการคำนวณค่าคุณลักษณะ Motion ในระดับของการหาเส้นทางการเดิน จะเป็นการพิจารณาความต่อเนื่องในแต่ละจุดข้อมูล แทนการมองภาพโดยรวมทั้งหมด

ค่าการเคลื่อนไหว (Motion-Cost) ในการคำนวณค่า Motion สำหรับคำตอบที่เป็นไปได้ V'_s จะใช้เวกเตอร์ความเร็วของแต่ละโหนดก่อนหน้ามาใช้ในการคำนวณ โดยการขยายออกมาอย่างต่อเนื่องเป็นช่วง

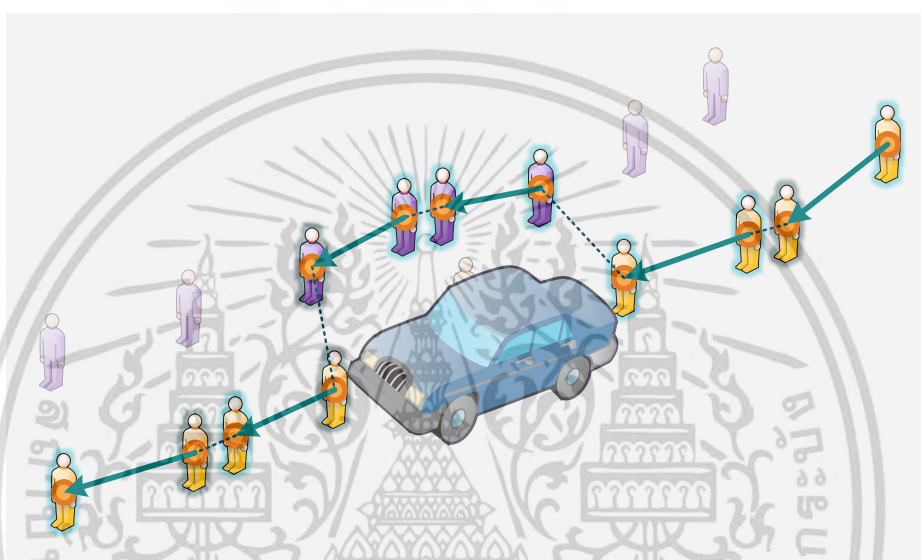
$$V'_{motion}(V'_s) = \sum_{i=3}^s |X'_s(i) - [X'_s(i-2) + 2\dot{X}'_s(i-2)]| \quad (4.18)$$

แต่อย่างไรก็ตาม ในขั้นตอนการเชื่อมโยงข้อมูลเพื่อสร้างเส้นทางการเดินของงานวิจัย [2] ผู้วิจัยพบว่าการใช้ค่าคุณลักษณะ Appearance มาใช้เพียงค่าเดียว และไม่นำโหนดสมมุติมาใช้พิจารณาร่วมด้วย จะให้ผลเส้นทางการเดินขั้นต้นที่เป็นไปได้ออกมาดีที่สุดในที่สุด แต่ก็ยังคงมีการติดตามสลับบุคคลเกิดขึ้นอยู่

4.5.2 แบ่งส่วนเส้นทางการเดินย่อยด้วยระยะห่างระหว่างแตร็คเล็ท (Trajectory Splitting)

หลังจากที่ได้ทำการสร้างเส้นทางการเดินขั้นต้น จากขั้นตอนก่อนหน้านี้ ในรูปที่ 4.12 เป็นตัวอย่างในกรณีที่มีการเลือกคำตอบสลับตัวบุคคล อันเนื่องมาจาก คนเดินถูกบดบังเป็นเวลานานทำให้เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไม่สามารถตรวจจับคนและส่งผลทำให้ไม่มีแทร็คเล็ทของตำแหน่งนั้นอยู่ ซึ่งคำตอบเส้นทางการเดิน เบื้องต้นที่มีลักษณะเลือกแทร็คเล็ทสลับตัวคน จะสังเกตได้ว่า ตำแหน่งบริเวณตรงที่เกิดการสลับตัวคน กันจะมีตำแหน่งการเดินทางที่ไม่ต่อเนื่องกัน ระยะห่างระหว่างแทร็คเล็ทตรงจุดที่ผิดพลาดจะมีค่ามากกว่าปกติ ด้วยแนวคิดนี้ งานวิจัยที่นำเสนอจึงใช้เกณฑ์ระยะห่างระหว่างแทร็คเล็ทเป็นเกณฑ์ในการแบ่ง ส่วนเส้นทางการเดินย่อย (Trajectory Segment) ด้วยสมมุติฐานที่ว่าแทร็คเล็ทที่อยู่ในส่วนย่อย เดียวกันและมีระยะห่างไม่มากนักมักจะเป็นแทร็คเล็ทที่เป็นของคนๆเดียวกัน

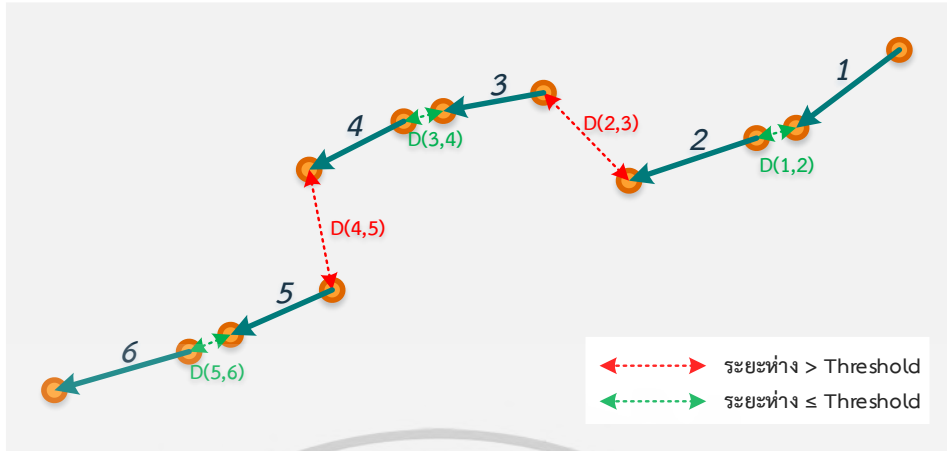


รูปที่ 4.12 กรณีการเลือกคำตอบของเส้นทางการเดินสลับตัวบุคคล

สำหรับขั้นตอนของการแบ่งส่วน Trajectory Segment โดยใช้ระยะห่างระหว่างแทร็คเล็ท ขั้นตอนแรก จะทำการคำนวณระยะห่างระหว่างแทร็คเล็ท โดยคิดจากจุดปลายแทร็คเล็ท ตำแหน่งปัจจุบัน (i) ไปยังจุดต้นของแทร็คเล็ทตำแหน่งถัดไป ($i + 1$) ตามสมการที่ 4.19

$$Dist(i, i + 1) = \sqrt{(x_{end\ point_i} - x_{start\ point_{i+1}})^2 + (y_{end\ point_i} - y_{start\ point_{i+1}})^2} \quad (4.19)$$

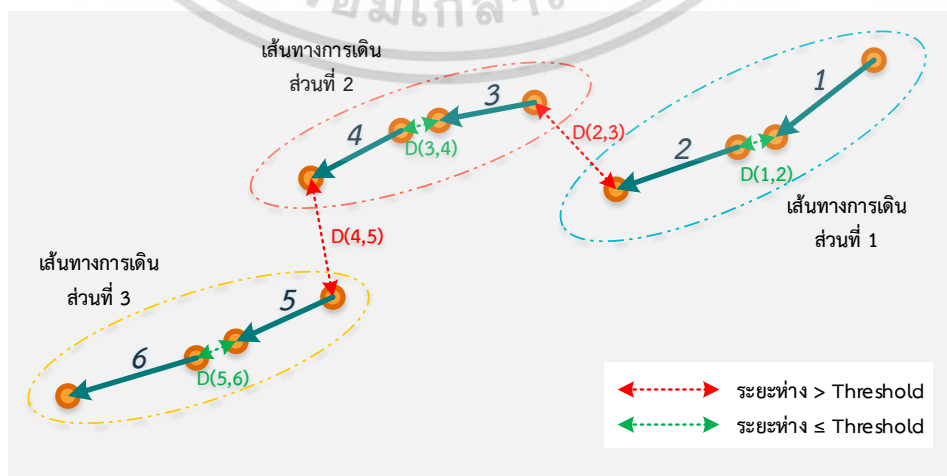
จากสมการที่ 4.19 กำหนดให้ i คือ ลำดับแทร็คเล็ทที่ i โดยที่ $1 \leq i \leq$ จำนวนแทร็คเล็ทของเส้นทางการเดินที่เป็นไปได้ $x_{start\ point}$, $y_{start\ point}$ คือ จุดพิกัด (x,y) เริ่มต้นแทร็คเล็ท $x_{end\ point}$, $y_{end\ point}$ คือ จุดพิกัด (x,y) สิ้นสุดของแทร็คเล็ท



รูปที่ 4.13 แสดงระยะห่างระหว่างแทร็คเล็ท

ขั้นตอนถัดไป คือ หาว่าระยะห่างระหว่างแทร็คเล็ทคู่ใดบ้างที่เกินค่าขีดแบ่ง โดยค่าขีดแบ่ง คือ ค่าที่แบ่งว่าแทร็คเล็ทคู่ใด มีความต่อเนื่องและมีโอกาสจะเป็นแทร็คเล็ทของบุคคลคนเดียวกัน จากรูปที่ 4.13 $D(2,3)$ และ $D(4,5)$ มีค่าที่มากกว่าค่าขีดแบ่ง จึงมีโอกาสดังที่แทร็คเล็ทที่ 2 เป็นคนละคนกับแทร็คเล็ทที่ 3 และแทร็คเล็ทที่ 4 มีโอกาสเป็นคนละคนกับแทร็คเล็ทที่ 5

ต่อมา จะสามารถพิจารณาเพื่อทำการแยก Trajectory segment ได้ โดยเงื่อนไขที่ใช้ในการแบ่งคือ ถ้าหากระยะห่างระหว่างแทร็คเล็ทคู่ใดๆ มีค่าน้อยกว่าค่าขีดแบ่ง จะกำหนดให้แทร็คเล็ทคู่นั้นเป็นส่วนเดียวกัน และถ้าหากระยะห่างระหว่างแทร็คเล็ทคู่ใดๆ มีค่ามากกว่าค่าขีดแบ่ง จะกำหนดให้แทร็คเล็ทคู่นั้นเป็นคนละส่วนกัน จากตัวอย่างรูปที่ 4.14 สามารถแยกส่วนของเส้นทางการเดินได้ออกเป็นสามส่วน แแทร็คเล็ทหมายเลข 1 และ 2 อยู่ในเส้นทางการเดินส่วนที่ 1 แแทร็คเล็ทหมายเลข 3 และ 4 อยู่ในเส้นทางการเดินส่วนที่ 2 และแทร็คเล็ทหมายเลข 5 และ 6 อยู่ในเส้นทางการเดินส่วนที่ 3

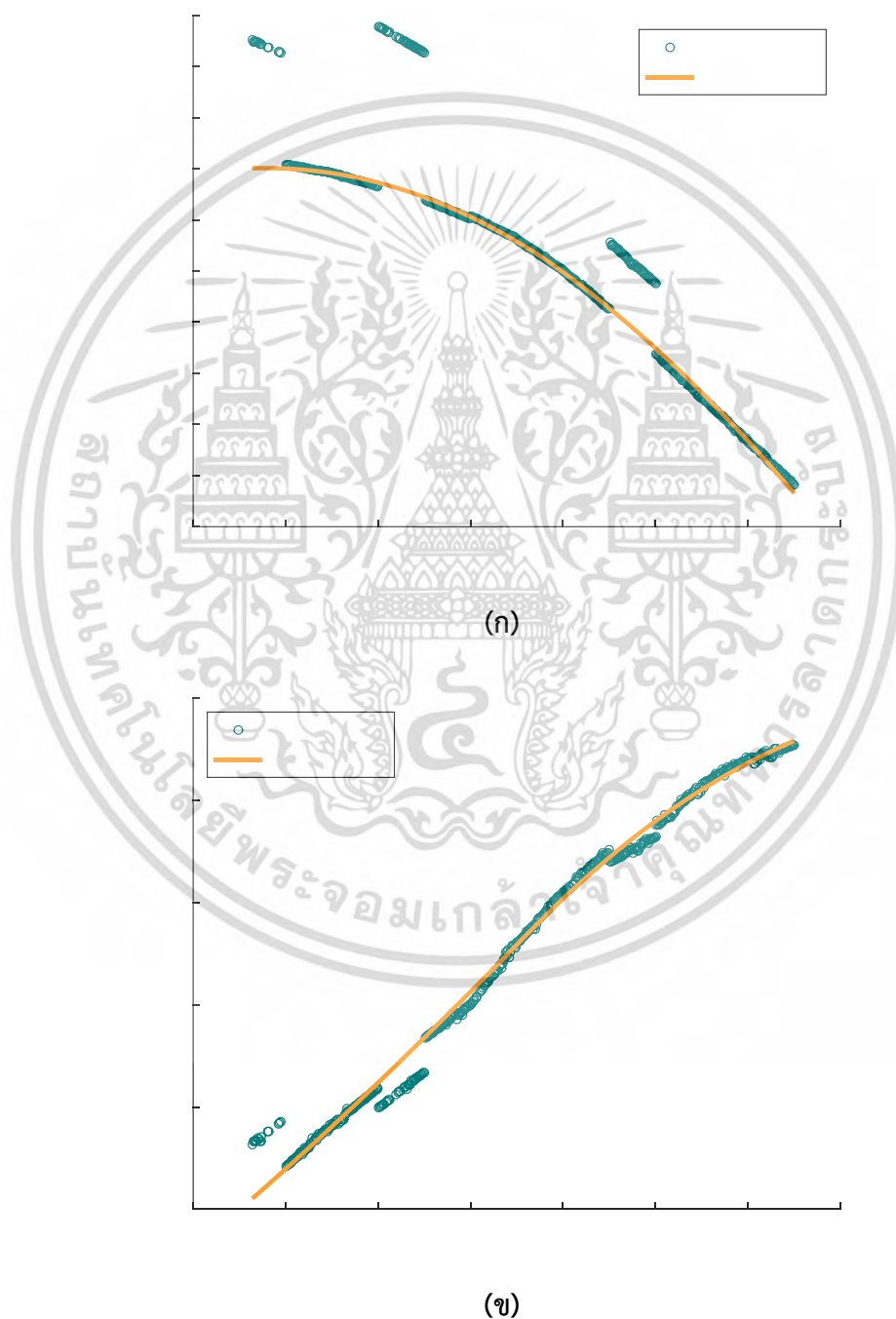


รูปที่ 4.14 การแบ่งส่วนย่อยของเส้นทางการเดิน

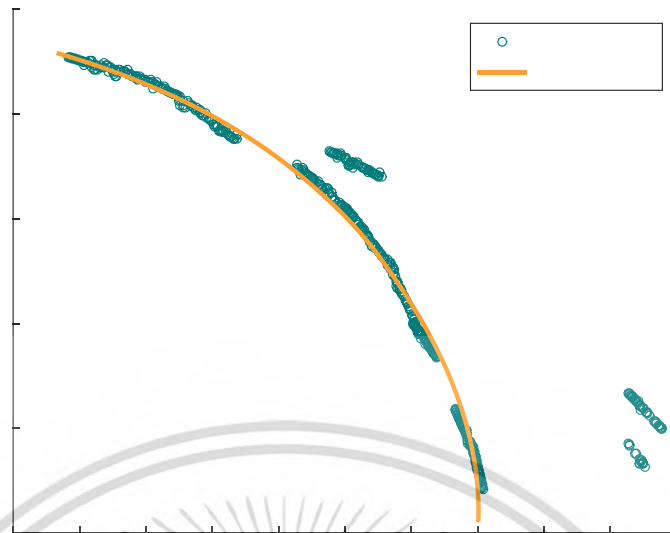
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.3 ประมาณเส้นทางการเดินด้วยเทคนิคการปรับเรียบข้อมูล (Trajectory Smoothing)

ในขั้นตอนนี้เป็นขั้นตอนที่ใช้เทคนิคการทำการปรับเรียบข้อมูล (smooth data) เพื่อให้รู้ว่าเส้นทางการเดินที่กำลังปรับปรุงอยู่ มีวิธีการเดินรูปแบบไหน โดยในขั้นตอนนี้จะแยกการทำการ trajectory smoothing ในแกน sequence(frame) - x และ sequence(frame) - y แยกกัน โดยใช้วิธีการ Robust Local Regression ในการทำ trajectory smoothing



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ค)

รูปที่ 4.15 การทำ smooth data บนแกนต่างๆ และเส้นประมาณเส้นทางการเดิน

(ก) การทำ smooth data บนแกน sequence (frame) - x

(ข) การทำ smooth data บนแกน sequence (frame) - y

(ค) เส้นที่ได้จากการทำ smooth data เทียบกับแทร็คเล็บบนแกน x-y

จากรูปที่ 4.16 เส้นสีส้มคือเส้นประมาณเส้นทางการเดินที่ได้จากการทำ smooth data จะเห็นว่า เส้นมีลักษณะที่มีแนวโน้มไปกับเส้นทางการเดินส่วนใหญ่ แม้เส้นทางการเดินที่วิเคราะห์จะมีแทร็คเล็ทที่ออกนอกแนววิถีไปจากแทร็คเล็ทอื่นๆก็ตาม



รูปที่ 4.16 เส้นประมาณเส้นทางการเดิน

กับแทร็คเล็ทภายในคำตอบเส้นทางการเดินขั้นต้นที่เป็นไปได้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์การเขียนเพื่อการศึกษาเท่านั้น เมื่อผู้ใดเห็นประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.4 การกำจัด trajectory segment ที่ผิดปกติ (Outlier Trajectory Segment Elimination)

เกณฑ์ที่ใช้ในการแบ่งระหว่างพิกัด inlier กับพิกัด outlier คือระยะห่างระหว่างพิกัดของเส้นทางการเดินกับพิกัดของเส้นประมาณเส้นทางการเดินที่ได้จากการทำการปรับเรียบข้อมูลในขั้นตอนที่ 4.5.3 โดยกำหนดว่า ถ้าระยะห่างมีค่ามากกว่าค่าขีดแบ่งให้ถือว่าพิกัดของเส้นทางการเดินดังกล่าวเป็นพิกัด outlier และถ้าหากน้อยกว่าหรือเท่ากับค่าขีดแบ่งให้เป็นพิกัด inlier จากหัวข้อที่ 4.5.2 ที่ได้ทำการแบ่ง Trajectory segment ด้วยระยะห่างระหว่างแทร็คเล็ตในตอนก่อนหน้านี้ ในขั้นตอนนี้จะทำการหาค่าเปอร์เซ็นต์ของพิกัด inlier ของแต่ละส่วนของ Trajectory segment

โดยที่ Trajectory segment ที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% กำหนดให้เป็น Trajectory segment ที่เป็น inlier สำหรับ Trajectory segment ที่มีค่าเปอร์เซ็นต์ของพิกัด inlier น้อยกว่า 50% จะถือว่าเป็น Trajectory segment ที่เป็น outlier และจะไม่ถูกนำไปใช้เป็นเส้นทางการเดินในรอบนั้น เนื่องจากมีแนวโน้มที่จะเป็นส่วนที่มีการติดตามสลับตัวบุคคลเกิดขึ้น



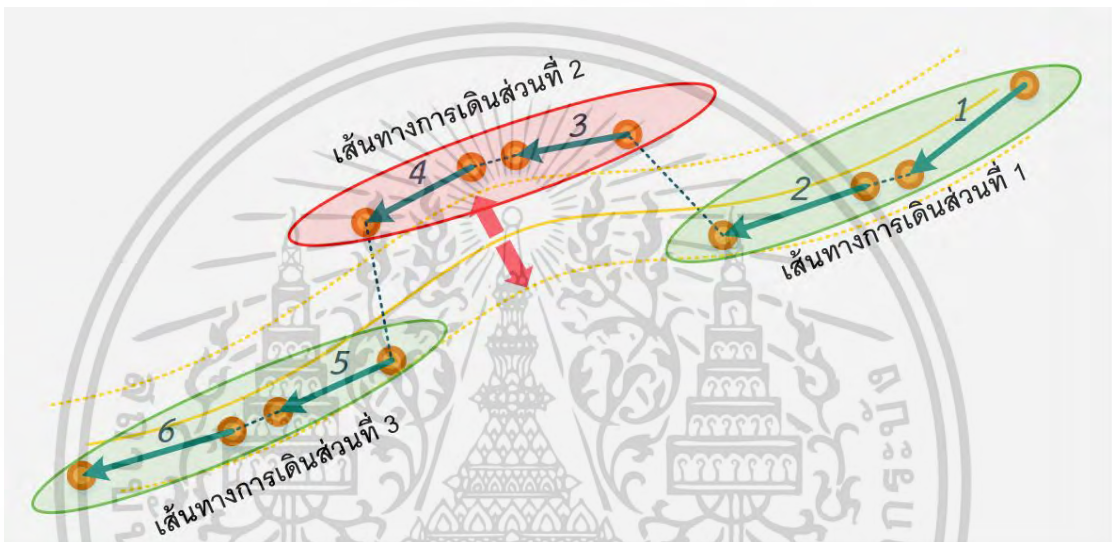
รูปที่ 4.17 กราฟแสดงขอบเขตในการแบ่งแยก inlier และ outlier

4.5.5 คัดเลือกส่วนของเส้นทางการเดินสมบูรณ์ของแต่ละคน (Trajectory Refinement)

เนื่องมาจาก Trajectory ที่มีเปอร์เซ็นต์ของพิกัด inlier ที่มีค่ามาก แสดงว่ามีตำแหน่งที่สอดคล้องกับวิถีเส้นทางการเดินโดยภาพรวม และมีโอกาสที่ส่วนดังกล่าวจะเป็นบุคคลคนเดียวกันกับ Trajectory segment อื่น เส้นทางการเดินที่มีการติดตามสลับคนบุคคล จะมีบางส่วนที่หลุดออกนอกวิถีไปเช่นในรูปที่ 4.18 จะเห็นว่า Trajectory segment ที่ 2 หลุดออกจากแนว ทำให้ค่าตอบของเส้นทางการเดินขาดความต่อเนื่อง เมื่อคิดค่าเปอร์เซ็นต์ของพิกัด inlier ของ Trajectory segment ทำให้มีค่าเปอร์เซ็นต์ที่น้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับขั้นตอนของการเลือกส่วนของเส้นทางการเดิน เพื่อเลือกมาเป็นคำตอบของเส้นทางการเดินของคน 1 คน จะเลือกจาก Trajectory segment ที่เป็น inlier สำหรับ Trajectory segment ที่เป็น outlier แทร็คเหล่านี้จะยังคงถูกเก็บไว้เป็นตัวเลือกในกราฟ G' สำหรับใช้ในการสร้างเส้นทางการเดินขั้นต้นที่เป็นไปได้ในรอบถัดๆไป ในกรณีที่ถ้าหากทุก Trajectory segment ไม่มีส่วนใดเลยที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% ในรอบนั้นจะเลือก Trajectory segment ที่มีจำนวนคำตอบหลายเฟรมที่สุด มาเป็นคำตอบสำหรับเส้นทางการเดินของรอบนั้น

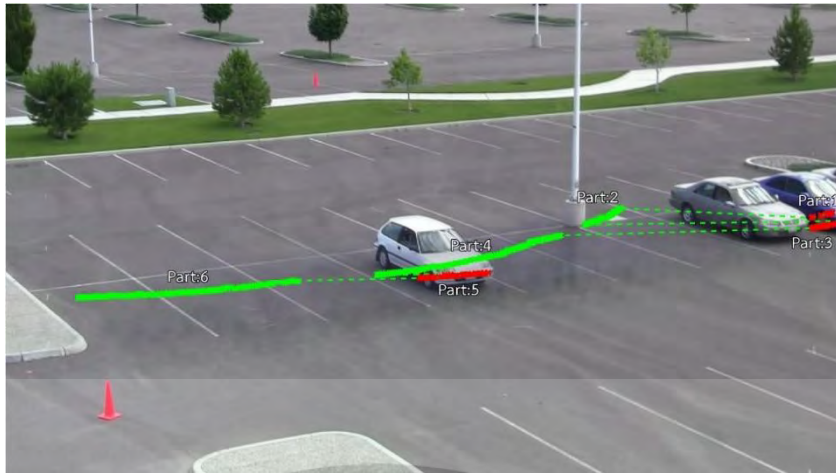


รูปที่ 4.18 ส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% (สีเขียว) และส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier น้อยกว่า 50% (สีแดง)

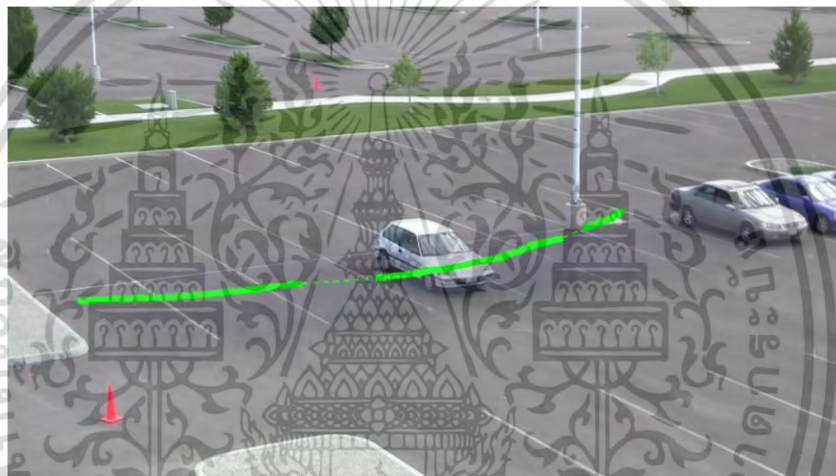


รูปที่ 4.19 คัดเฉพาะส่วนย่อยเส้นทางการเดินที่มีเปอร์เซ็นต์ของพิกัด inlier มากกว่าหรือเท่ากับ 50% เพื่อมาเป็นคำตอบเส้นทางการเดินของคน 1 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 พิกัดที่เป็นพิกัด inlier (สีเขียว) และพิกัด outlier (สีแดง)



(ก)



(ข)

รูปที่ 4.21 เส้นทางการเดินของคน 1 คน

(ก) ผลเส้นทางการเดินจากชุดการทดลอง Parking Lot

(ข) ผลเส้นทางการเดินจากชุดการทดลอง Town Centre

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

การทดลองและผลการทดลอง

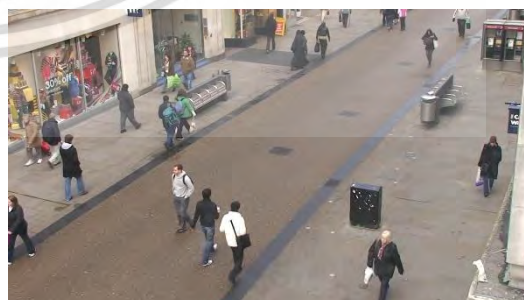
ในบทนี้จะกล่าวถึงการทดลองและผลการทดลอง เพื่อการวัดประสิทธิภาพของการติดตามวัตถุของงานวิจัยที่นำเสนอ นอกจากนี้ในบทนี้ยังอธิบายถึงการทดลองสำหรับการวิเคราะห์ค่าตัวแปรที่เหมาะสมในแต่ละขั้นตอน โดยเริ่มแรกจากขั้นตอนการตรวจจับบุคคล หัวข้อที่ 5.2 จะวิเคราะห์หาค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมสำหรับการขจัดค่าที่ไม่มากที่สุด (Nonmaximum Suppression) เพื่อให้ได้ผลการตรวจจับบุคคลที่มีประสิทธิภาพมากที่สุด ขั้นตอนของการติดตามบุคคลมีขั้นตอนที่จะต้องปรับปรุงเส้นทางการเดินด้วยเทคนิคการปรับเรียบข้อมูล (smooth data) และคำนวณแต่ละส่วนของเส้นทางการเดิน (Trajectory segment) ที่มีพิทักอยู่ในเกณฑ์ (inlier) สำหรับเลือกมาเป็นผลลัพธ์เส้นทางการเดิน หัวข้อที่ 5.3 จะอธิบายถึงการวิเคราะห์ค่าขีดแบ่งสำหรับการหาค่าที่อยู่ในเกณฑ์ (inlier) และช่วงขอบเขตข้อมูลที่ใช้การประมาณเส้นทางการเดิน และในหัวข้อที่ 5.4 จะนำเสนอผลการทดลองวัดประสิทธิภาพของการติดตามวัตถุของงานวิจัยที่นำเสนอเทียบกับวิธีการ GMCP-Tracker[2] ในแง่มุมต่างๆ

5.1 ชุดข้อมูล

สำหรับชุดการทดลองที่งานวิจัยที่นำเสนอใช้ในการวัดประสิทธิภาพนั้น ใช้ชุดการทดลอง Parking Lot จากงานวิจัย [2] ลักษณะภาพเป็นภาพที่มีฝูงคนจำนวน 14 คนเดินอยู่ในลานจอดรถพร้อมกัน แต่ละคนเดินมีลักษณะสีเสื้อการแต่งกายที่ใกล้เคียงกัน และชุดการทดลอง Town Centre มีลักษณะที่คนเดินสวนทางกัน และคนเดินเข้ามาในเฟรมภาพไม่พร้อมกัน ซึ่งทั้งสองชุดการทดลองแต่ละภาพมีขนาด 1920 x 1080 พิกเซล มีเหตุการณ์ที่คนเดินบดบังกันเองบ่อยครั้งและคนเดินถูกบดบังจากวัตถุ



(ก)



(ข)

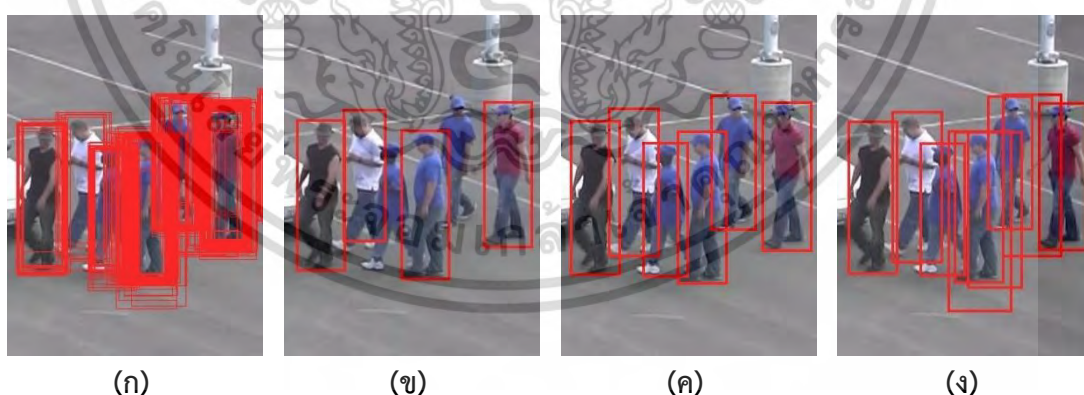
รูปที่ 5.1 ชุดการทดลอง

(ก) ชุดการทดลอง Parking Lot (ข) ชุดการทดลอง Town Centre

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 การวิเคราะห์ค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมสำหรับการขจัดค่าที่ไม่มากที่สุด (Nonmaximum Suppression) เพื่อเพิ่มประสิทธิภาพการตรวจจับบุคคลจากภาพนิ่ง

ในขั้นตอนแรกของงานวิจัยที่นำเสนอจะต้องทำการตรวจจับบุคคลจากภาพนิ่งของวิดีโอ โดยเลือกใช้วิธีการจากงานวิจัย [1] มาใช้ในการตรวจจับบุคคล ซึ่งมีขั้นตอนการทำ image pyramid จึงทำให้มีผลการตรวจจับวัตถุขึ้นเดียวกันในหลากหลายขนาดภาพ ขั้นตอนต่อมาได้ทำการขจัดค่าที่ไม่มากที่สุด (Nonmaximum Suppression) โดยขั้นตอนนี้เป็นการเลือกผลการตรวจจับที่เหมาะสมที่สุดจากผลการตรวจจับวัตถุหลากหลายขนาด เพื่อลดความซ้ำซ้อนของผลการตรวจจับ โดยจะพิจารณาผ่านพื้นที่ทับซ้อนระหว่างกรอบผลการตรวจจับ เช่น ถ้ามีผลการตรวจจับที่มีค่าคะแนนที่ดีกว่าผลการตรวจจับอีกผลหนึ่ง ถ้าผลการตรวจจับทั้งสองมีพื้นที่ทับซ้อนกันมากกว่าค่าขีดแบ่งที่กำหนดไว้จะถือว่าผลการตรวจจับทั้งคู่เป็นผลการตรวจจับของวัตถุขึ้นเดียวกัน ดังนั้นจะเลือกผลการตรวจจับที่มีค่าคะแนนที่ดีกว่าเป็นคำตอบเพียงคำตอบเดียว ในหัวข้อนี้จะเป็นการวิเคราะห์พารามิเตอร์ค่าขีดแบ่งพื้นที่ทับซ้อนที่ใช้ในการทำ Nonmaximum Suppression เพื่อศึกษาว่าการกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนมีผลต่อแนวโน้มค่าผลบวกจริง (TP) ผลบวกлож (FP) ผลลบлож (FN) อย่างไรและศึกษาการเลือกกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสม โดยพิจารณาจากค่าความถูกต้องในการตรวจจับหลายวัตถุ (N-MODA) และค่าความแม่นยำในการตรวจจับหลายวัตถุ (N-MODP) เพื่อให้ได้ประสิทธิภาพการตรวจจับที่ดีที่สุด เพื่อนำการตรวจจับไปใช้ในการเชื่อมโยงข้อมูลเพื่อสร้างแทร็คเล็ท ในขั้นตอนถัดไป



รูปที่ 5.2 การทำการขจัดค่าที่ไม่มากที่สุดโดยการใช้ค่าขีดแบ่งพื้นที่ทับซ้อนต่างๆ

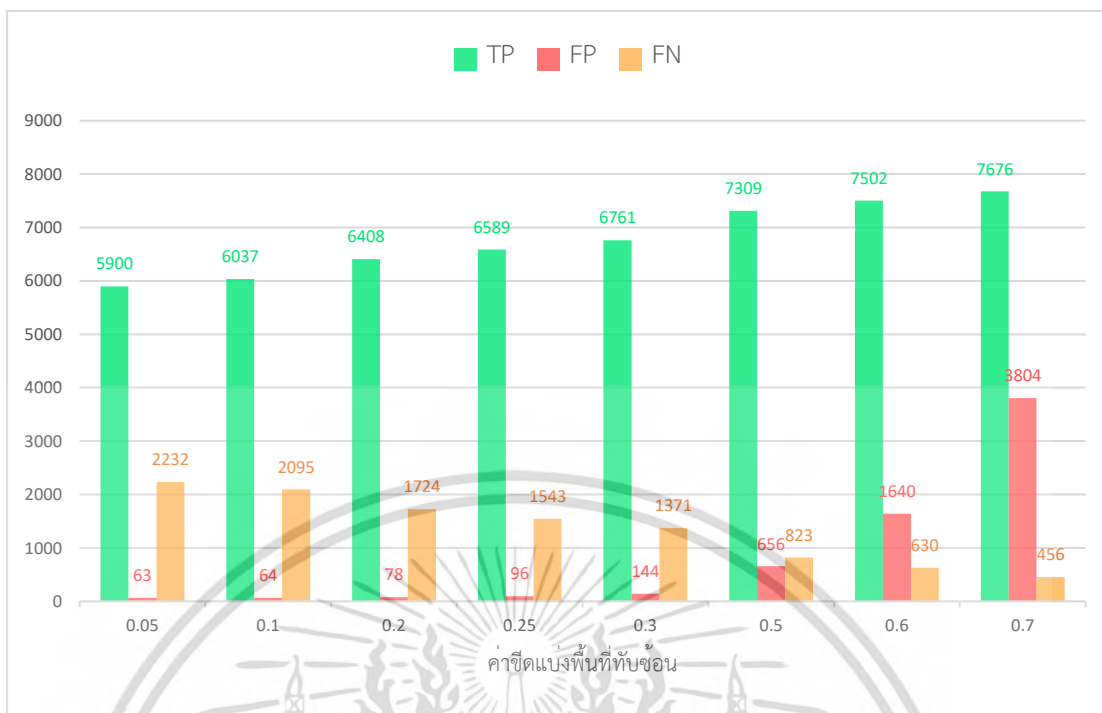
(ก) ผลการตรวจจับก่อนทำการขจัดค่าที่ไม่มากที่สุด

(ข) ค่าขีดแบ่งพื้นที่ทับซ้อน = 0.05

(ค) ค่าขีดแบ่งพื้นที่ทับซ้อน = 0.3

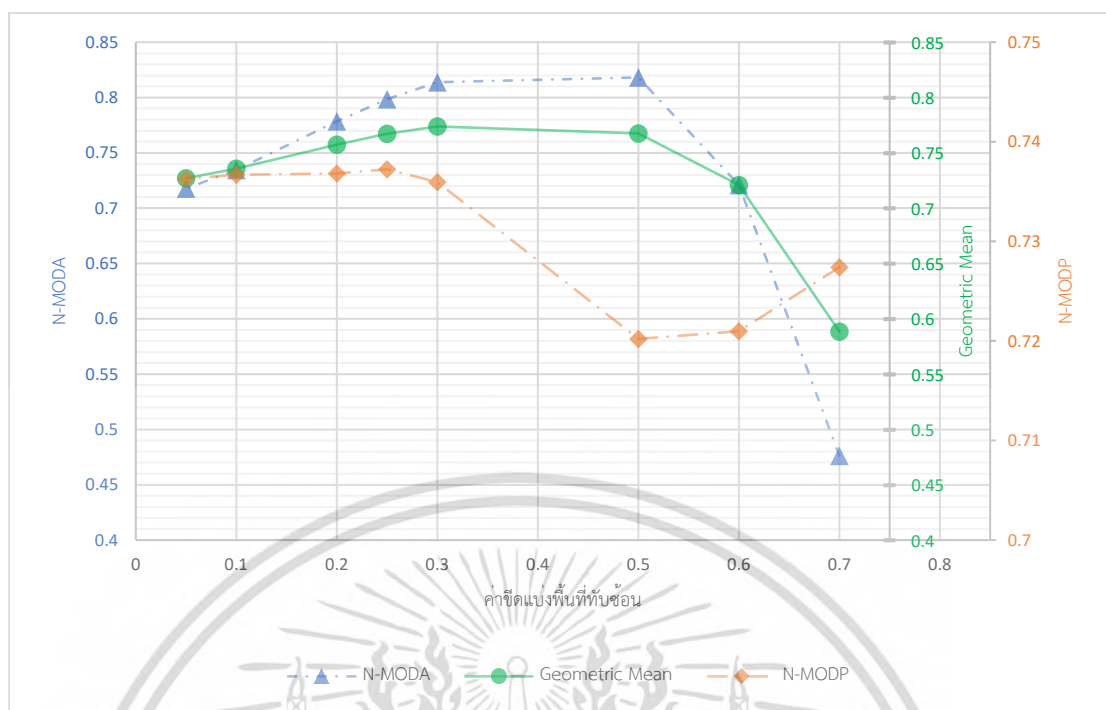
(ง) ค่าขีดแบ่งพื้นที่ทับซ้อน = 0.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.3 กราฟแสดงความสัมพันธ์ระหว่างค่าขีดแบ่งพื้นที่ทับซ้อน กับ TP FP และ FN

การวัดประสิทธิภาพการตรวจจับด้วยค่า TP FP และ FN เป็นมาตรวัดพื้นฐานที่จะนำไปใช้ในการคำนวณในการวัดประสิทธิภาพการตรวจจับในรูปแบบอื่นได้อีก จากรูปที่ 5.3 ผลการทดลองพบว่า การกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่ค่ามากเกินไป จะส่งผลให้ค่า TP ยังมีค่ามาก แต่ก็ต้องแลกมาด้วยค่า FP ที่มากขึ้นเช่นกัน เพราะว่าการกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่มีค่ามาก ทำให้ผลลัพธ์การตรวจจับจะต้องมีพื้นที่ทับซ้อนระหว่างกันที่มากกว่าค่าขีดแบ่ง จึงจะทำให้ผลการตรวจจับที่มีค่าคะแนนที่น้อยกว่าไม่ถูกเลือกเป็นคำตอบ ถ้าผลการตรวจจับมีพื้นที่ทับซ้อนกันน้อยกว่าค่าขีดแบ่ง ผลการตรวจจับที่ค่าคะแนนน้อยกว่าย่อมจะถูกเลือกไปเป็นคำตอบด้วย แม้จะมีผลการตรวจจับที่มีค่าคะแนนที่มากกว่าถูกเลือกเป็นผลลัพธ์อยู่แล้วก็ตาม ซึ่งกรณีนี้ส่งผลให้มีผลลัพธ์ของการตรวจจับวัตถุมีคำตอบที่ซ้ำซ้อน ดังตัวอย่างในรูปที่ 5.2 (ง) ส่งผลให้ค่า FP มีค่าสูงมากยิ่งขึ้น สำหรับการกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่มีค่าน้อยเกินไป พบว่า FP มีค่าน้อยลง แต่ก็ต้องแลกมาด้วยค่า TP ที่น้อยลงด้วยเช่นกัน เพราะว่าการใช้ค่าขีดแบ่งพื้นที่ทับซ้อนน้อยจะสามารถลดความซ้ำซ้อนของผลลัพธ์ที่เป็นวัตถุขึ้นเดียวกันได้ดี แต่ในกรณีที่คนมีตำแหน่งใกล้เคียงกันมาก จนทำให้เกิดพื้นที่ทับซ้อนระหว่างกัน การที่กำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่มีค่าน้อยเกินไป อาจทำให้เกิดพื้นที่ทับซ้อนระหว่างผลการตรวจจับของคนที่มีตำแหน่งใกล้เคียงกันมีค่ามากกว่าค่าขีดแบ่งพื้นที่ทับซ้อน ส่งผลให้ผลการตรวจจับของคนใดคนหนึ่งตำแหน่งใกล้เคียงกันไม่ถูกเลือกเป็นคำตอบ ดังตัวอย่างในรูปที่ 5.2 (ข) จากกรณีนี้จึงมีผลทำให้ค่า TP น้อยลงไปด้วยเช่นกัน



รูปที่ 5.4 กราฟแสดงความสัมพันธ์ระหว่างค่าขีดแบ่งพื้นที่ทับซ้อนกับ N-MODA N-MODP และ Geometric Mean

ค่าความถูกต้องในการตรวจจับหลายวัตถุ (N-MODA) เป็นการวัดที่ใช้ค่า FP และ FN มาเป็นเกณฑ์ในการคำนวณ การกำหนดค่าขีดแบ่งพื้นที่ทับซ้อน ที่ให้ FP และ FN ที่น้อยจะทำให้ ค่าความถูกต้องในการตรวจจับหลายวัตถุมากขึ้น ยิ่งมีค่ามากแสดงว่ายังมีประสิทธิภาพการตรวจจับที่ดี ค่าความแม่นยำในการตรวจจับหลายวัตถุ (N-MODP) สามารถใช้เป็นเกณฑ์ในการวัดว่าผลรอบสี่เหลี่ยมการตรวจจับมีพื้นที่ทับซ้อนตรงกับวัตถุมากน้อยเพียงใด ยิ่งมีค่ามากแสดงว่าผลการตรวจจับยังมีความพอดีกับตำแหน่งวัตถุมาก ซึ่งทั้งสองค่านี้ เป็นเกณฑ์ที่มีความสำคัญในการพิจารณาเลือกค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมที่สุดด้วย รูปที่ 5.4 จากผลการทดลอง พบว่าการกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.6 ขึ้นไป ค่าความถูกต้องในการตรวจจับหลายวัตถุมีค่าที่แยกลง เพราะว่าถึงแม้ค่า TP มีแนวโน้มที่มากขึ้นก็ตาม แต่ค่า FP ก็เพิ่มขึ้นมากเช่นกัน เช่น ในจุดค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.5 เทียบกับ 0.6 พบว่า ค่า FP เพิ่มขึ้นถึง 150% ในขณะที่ค่า FN กลับลดลงแค่เพียง 23.45% ซึ่งหมายถึงการมีคำตอบที่ซ้ำซ้อนมากยิ่งขึ้นนั่นเอง กรณีนี้จึงส่งผลให้ค่าความถูกต้องในการตรวจจับหลายวัตถุมีค่าน้อยลง ในช่วงที่กำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่ค่ามาก ค่าความแม่นยำในการตรวจจับหลายวัตถุมีค่าน้อยลงกว่าช่วงกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่ค่าน้อยแค่เพียงเล็กน้อย สำหรับค่าความถูกต้องในการตรวจจับหลายวัตถุที่ดีที่สุดจะอยู่ตรงตำแหน่งที่กำหนดค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.5 และค่าความถูกต้องในการตรวจจับหลายวัตถุที่ตรงลงมาอยู่ที่ตำแหน่งที่กำหนดค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.3 แต่ค่าความแม่นยำในการตรวจจับหลายวัตถุ ค่าขีดแบ่งเท่ากับ 0.5 มีค่าน้อยกว่าค่าขีดแบ่งเท่ากับ 0.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับเกณฑ์ที่จะใช้เลือกค่าขีดแบ่งพื้นที่ที่ทับซ้อนที่เหมาะสมที่สุด งานวิจัยที่นำเสนอใช้ ค่าเฉลี่ยเรขาคณิต (Geometric Mean) ดังในสมการที่ 3.36 โดยพิจารณาจากค่าความถูกต้องในการตรวจจับหลายวัตถุ และค่าความแม่นยำในการตรวจจับหลายวัตถุไปพร้อมกัน เพื่อเป็นเกณฑ์ในการพิจารณาค่าขีดแบ่งพื้นที่ที่ทับซ้อนที่เหมาะสมที่สุด เป็นการสร้าง สมดุลระหว่างค่าทั้งสอง จากผลการทดลองพบว่าค่าขีดแบ่งพื้นที่ที่ทับซ้อนเท่ากับ 0.3 เป็นค่าที่ทำให้ค่าเฉลี่ยเรขาคณิตมีค่ามากที่สุด ดังนั้นค่าขีดแบ่งพื้นที่ที่ทับซ้อนเท่ากับ 0.3 จะเป็นค่าที่เหมาะสมที่สุดในการหา Nonmaximum Suppression สำหรับการทดลองครั้งนี้

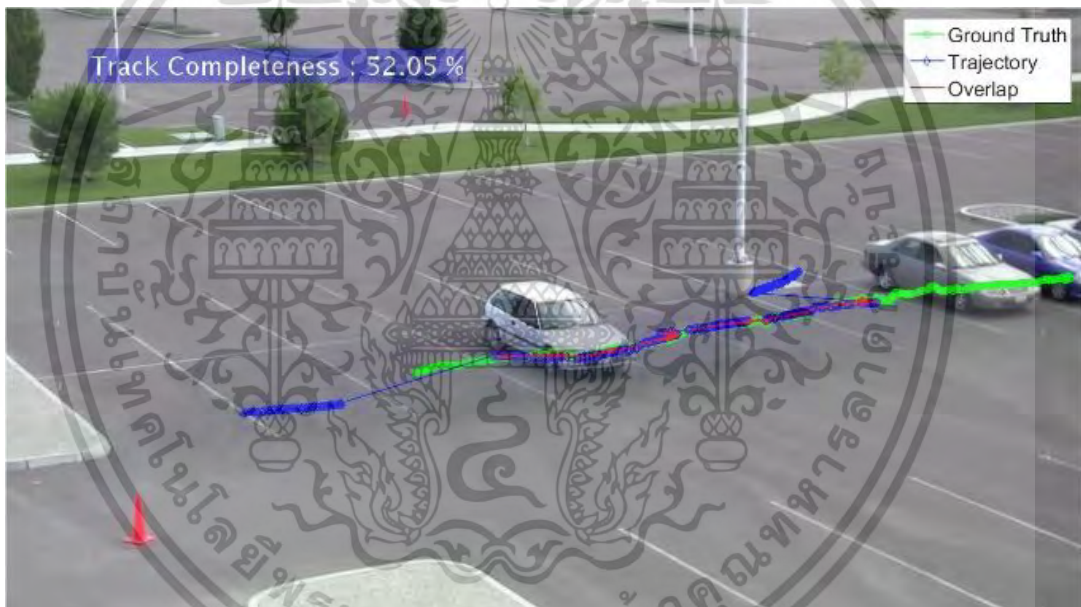
5.3 การวิเคราะห์ค่าขีดแบ่งสำหรับการหาค่าที่อยู่ในเกณฑ์และช่วงขอบเขตข้อมูลที่ใช้ในการประมาณเส้นทางการเดิน

จากหัวข้อที่ 4.5.3 ที่อธิบายถึงการคำนวณพิกัดของเส้นประมาณเส้นทางการเดินด้วยเทคนิคการปรับเรียบข้อมูล (smooth data) เพื่อหาวิถีการเดินโดยประมาณของเส้นทางการเดินที่กำลังปรับปรุงอยู่ ซึ่งในขั้นตอนดังกล่าว ใช้วิธีการ Robust Local Regression ในการคำนวณ ซึ่งจะต้องมีการกำหนดขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดิน และหัวข้อที่ 4.5.4 ที่อธิบายถึงขั้นตอนของการคำนวณจำนวนพิกัดที่อยู่ในเกณฑ์ ของแต่ละส่วนของเส้นทางการเดิน (Trajectory segment) ซึ่งในขั้นตอนดังกล่าว จะต้องมีการกำหนดค่าขีดแบ่งของระยะห่างระหว่างพิกัดของเส้นทางการเดินกับพิกัดของเส้นประมาณเส้นทางการเดินที่ได้จากการปรับเรียบข้อมูล เพื่อแบ่งว่าข้อมูลพิกัดจุดใดเป็นพิกัดที่อยู่ในเกณฑ์ (inlier) หรือพิกัดที่อยู่นอกเกณฑ์ (outlier) ดังนั้นในหัวข้อนี้ จะเป็นการทดลองเพื่อวิเคราะห์หาพารามิเตอร์ที่เหมาะสมของค่าขอบเขตช่วงข้อมูลที่ใช้การประมาณเส้นทางการเดิน และค่าขีดแบ่งระยะห่างในการเลือกพิกัด inlier

โดยการวัดประสิทธิภาพที่ใช้พิจารณาในหัวข้อนี้ จะใช้ค่าความสมบูรณ์ของการติดตาม (Track Completeness) จากในสมการที่ 3.30 โดยคำนวณจากจำนวนพิกัดที่คาบเกี่ยวกันระหว่างเส้นทางการเดินที่ถูกต้อง (Ground Truth) กับผลลัพธ์เส้นทางการเดิน โดยพิจารณาเฟรมต่อเฟรมเทียบกับจำนวนพิกัดทั้งหมดของเส้นทางการเดินที่ถูกต้อง โดยยังมีค่ามากจะยิ่งถือว่าผลลัพธ์เส้นทางการเดินมีตำแหน่งที่ใกล้กับเส้นทางการเดินที่ถูกต้องและมีประสิทธิภาพการติดตามที่ดี จากรูปที่ 5.5 (ก) ค่าสมบูรณ์ของการติดตามมีค่าเท่ากับ 99.68% เป็นตัวอย่างของผลลัพธ์เส้นทางการเดินที่มีความถูกต้องสูง จะเห็นว่าผลลัพธ์เส้นทางการเดินกับเส้นทางการเดินที่ถูกต้องมีพิกัดตำแหน่งที่ใกล้กันและซ้อนทับกันตลอดเส้นทางการเดิน จากรูปที่ 5.5 (ข) ค่าสมบูรณ์ของการติดตามมีค่าเท่ากับ 52.05% เป็นตัวอย่างของผลลัพธ์เส้นทางการเดินที่มีความผิดพลาด จะเห็นว่าผลลัพธ์เส้นทางการเดินมีตำแหน่งที่ขาดหายและมีการติดตามที่เลือกตำแหน่งของบุคคลอื่นมาเป็นคำตอบ ทำให้เกิดระยะห่างออกไปจากตำแหน่งเส้นทางการเดินที่ถูกต้องของบุคคลที่ติดตามอยู่



(ก)



(ข)

รูปที่ 5.5 ผลลัพธ์เส้นทางการเดินเทียบกับเส้นทางการเดินที่ถูกต้อง

(ก) ผลลัพธ์เส้นทางการเดินที่วัดค่าความสมบูรณ์การติดตามเท่ากับ 99.68%

(ข) ผลลัพธ์เส้นทางการเดินที่วัดค่าความสมบูรณ์การติดตามเท่ากับ 52.05%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 ค่าความสมบูรณ์ของการติดตามของเส้นทางการเดินที่เลือกใช้วิธีการปรับเรียบ
ข้อมูล Robust Local Regression และ Savitzky-Golay

	วิธีการ Smooth Data	
	Robust Local Regression	Savitzky-Golay
ค่าความสมบูรณ์ของการติดตาม	76.56%	75.41%

จากหัวข้อที่ 4.5.3 ที่อธิบายถึงการคำนวณพิกัดของเส้นประมาณเส้นทางการเดินด้วยเทคนิคการปรับเรียบข้อมูล (smooth data) เพื่อหาวิถีการเดินโดยประมาณของเส้นทางการเดินที่กำลังปรับปรุงอยู่ จากการทดลองพบว่าการเลือกใช้วิธีการปรับเรียบข้อมูลวิธี Robust Local Regression ให้ค่าความสมบูรณ์ของการติดตามเท่ากับ 76.56% ซึ่งมีค่ามากกว่าการใช้วิธีการปรับเรียบข้อมูลวิธี Savitzky-Golay ซึ่งให้ค่าความสมบูรณ์ของการติดตามเท่ากับ 75.41% ดังนั้นงานวิจัยที่นำเสนอ จึงเลือกใช้วิธีการ Robust Local Regression ในการปรับเรียบข้อมูล สำหรับขั้นตอนในหัวข้อที่ 4.5.3

ในหัวข้อนี้ ได้ทดลองทำการปรับค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินในช่วง 0.1 ถึง 0.9 ทำการปรับค่าขีดแบ่งระยะห่างในการเลือกพิกัด inlier เท่ากับ 5 10 และ 15 เพื่อศึกษาว่าค่าช่วงใดที่เป็นค่าที่เหมาะสม ที่ส่งผลให้ได้ค่าความสมบูรณ์ของการติดตามที่มากที่สุด จากผลทดลองในตารางที่ 5.2 การกำหนดค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินเท่ากับ 0.2 ให้ผลลัพธ์เส้นทางการเดินที่มีค่าความสมบูรณ์ของการติดตามมากที่สุด จากผลทดลอง พบว่าค่าความสมบูรณ์ของการติดตามที่ดีที่สุดเท่ากับ 81.06% ซึ่งเป็นผลลัพธ์ที่ได้จากการกำหนดค่าขีดแบ่งระยะห่างในการเลือกค่า inlier เท่ากับ 5 และค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินเท่ากับ 0.2

ตารางที่ 5.2 ค่าความสมบูรณ์ของการติดตามของเส้นทางการเดินที่ได้จากการปรับค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินและค่าขีดแบ่งระยะห่างในการเลือกค่าที่อยู่ในเกณฑ์

		ค่าขีดแบ่งระยะห่าง ในการเลือกค่าที่อยู่ในเกณฑ์		
		5	10	15
ค่าขอบเขตช่วงข้อมูลที่ใช้การ ประมาณเส้นทางการเดิน	0.1	75.49%	76.46%	74.12%
	0.2	81.06%	77.19%	74.98%
	0.3	76.71%	78.41%	75.54%
	0.4	77.97%	76.67%	76.46%

ตารางที่ 5.2 (ต่อ)

	0.5	75.91%	78.91%	77.26%
	0.6	75.96%	78.73%	78.36%
	0.7	72.80%	77.43%	78.39%
	0.8	74.77%	76.18%	78.21%
	0.9	70.76%	74.91%	77.55%

5.4 การวัดประสิทธิภาพของการติดตามวัตถุเทียบกับวิธีการ GMCP-Tracker[2]

ในหัวข้อที่ 5.4 จะนำเสนอผลการทดลองเปรียบเทียบประสิทธิภาพการติดตามวัตถุของงานวิจัยที่นำเสนอกับงานวิจัย GMCP-Tracker[2] เพื่อแสดงให้เห็นผลลัพธ์จากงานวิจัยที่นำเสนอที่สามารถช่วยปรับปรุงเส้นทางการเดินให้มีผลลัพธ์ที่มีประสิทธิภาพที่ดีขึ้นและได้ลดปัญหาการติดตามคนเดินสลับบุคคล เกณฑ์ชี้วัดในหัวข้อนี้จะแบ่งเกณฑ์การพิจารณาเป็น 2 รูปแบบ ได้แก่ แบบพิจารณาจากเส้นทางการเดินทั้งหมดตลอดทั้งวิดีโอ กับแบบพิจารณาผลลัพธ์การติดตามแบบเฟรมต่อเฟรม โดยสำหรับแบบพิจารณาจากเส้นทางการเดินทั้งหมดตลอดทั้งวิดีโอจะใช้ตัวชี้วัดคือ ค่าความสมบูรณ์ของการติดตาม (Track Completeness) และค่าระยะที่ผิดพลาดของการติดตาม (Track Matching Error : TME) โดยค่าความสมบูรณ์ของการติดตามเป็นการพิจารณาระหว่างเส้นทางการเดินที่ถูกต้องกับผลลัพธ์เส้นทางการเดิน เป็นการวัดว่าพิกัดคำตอบมีระยะใกล้เคียงกับเส้นทางการเดินที่ถูกต้องและจำนวนพิกัดคำตอบเป็นของคนเดียวกันตลอดทั้งเส้นทางการเดินมากน้อยเพียงใด ในส่วนของค่าระยะที่ผิดพลาดของการติดตาม เป็นการคำนวณระยะห่างเฉลี่ยระหว่างพิกัดเส้นทางการเดินที่ถูกต้องกับพิกัดผลลัพธ์เส้นทางการเดิน สำหรับการวัดประสิทธิภาพแบบพิจารณาผลลัพธ์การติดตามแบบเฟรมต่อเฟรมจะใช้ตัวชี้วัดคือ ค่าความถูกต้องในการติดตามหลายวัตถุ (MOTA) และจำนวนการครั้งการสลับตัวบุคคล (ID-Switches) โดยค่าความถูกต้องในการติดตามหลายวัตถุ เป็นการนำค่าผลบวกหลง (FP) ค่าผลลบหลง (FN) และจำนวนการครั้งการสลับตัวบุคคล มาคำนวณเพื่อหาค่าความถูกต้อง

ตารางที่ 5.3 การเปรียบเทียบประสิทธิภาพของงานวิจัยที่นำเสนอกับวิธี GMCP[2]

ชุดการทดลอง	งานวิจัย	ค่าความสมบูรณ์ของการติดตาม	TME	MOTA	ID-Switches
Parking Lot	GMCP-Tracker[2]	72.89%	39.09	86.16%	25
	งานวิจัยที่นำเสนอ	81.06%	16.17	86.41%	10
Town Centre	GMCP-Tracker[2]	75.61%	129.75	71.53%	19
	งานวิจัยที่นำเสนอ	79.38%	82.80	70.69%	13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานานับ ไม่สามารถตีพิมพ์ไปเผยแพร่โดยไม่ขออนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 5.3 ผลการทดลองพบว่า การเปรียบเทียบประสิทธิภาพด้วยการวัดค่าความถูกต้องในการติดตามหลายวัตถุ (MOTA) พบว่าประสิทธิภาพของงานวิจัยที่นำเสนอและวิธี GMCP-Tracker ไม่แตกต่างกันมาก อันเนื่องมาจากค่าที่วัดได้เท่า การนับจำนวนครั้งที่เกิดการติดตามบุคคลสลับตัวในแต่ละเฟรมจะนับแค่เพียงครั้งที่เกิดการระบุตัวบุคคลแตกต่างกับเฟรมก่อนหน้าเพียงครั้งเดียว หลังจากนั้นในการติดตามต่อไปถ้าหากการระบุตัวบุคคลเป็นคนเดียวกับเฟรมก่อนหน้าจะถือว่า การติดตามนั้นไม่เกิดการสลับเกิดขึ้น ถึงแม้ผลลัพธ์จะติดตามผิดคน โดยคนนั้นจะไม่ใช้คนเดียวกันกับที่ผลลัพธ์เส้นทางการเดินส่วนใหญ่ที่กำลังติดตามอยู่ก็ตาม และนอกจากนี้การคำนวณค่าความถูกต้องในการติดตามหลายวัตถุ ค่า FP FN และจำนวนการครั้งการสลับตัวบุคคล มาคิดเป็นความผิดพลาดรวมกันแล้วหารด้วยจำนวนพิกัด Ground Truth ทั้งหมด ดังนั้นถึงแม้ปริมาณผลการสลับตัวบุคคลของงานวิจัยที่นำเสนอจะให้ค่าลดลง ไม่ได้ส่งผลที่แตกต่างมากเท่าไร

สำหรับการวัดประสิทธิภาพแบบพิจารณาจากเส้นทางการเดินทั้งหมดตลอดทั้งวิดีโอ ในก่อนที่จะทำการวัดค่าความสมบูรณ์ของการติดตามและค่าระยะที่ผิดพลาดของการติดตาม จะต้องทำการจับคู่ระหว่างผลลัพธ์เส้นทางการเดินกับคำตอบเส้นทางการเดินที่ถูกต้องก่อน เพื่อที่จะทราบว่าผลลัพธ์เส้นทางการเดินโดยพิจารณาจากตลอดทั้งวิดีโอแต่ละคำตอบกำลังติดตามบุคคลคนใดอยู่ โดยจับคู่ภายใต้เงื่อนไขที่ว่าคู่ใดมีจำนวนพิกัดใกล้เคียงกันมากที่สุดก็จะนับว่าเป็นคู่ที่เหมาะสมกัน ต่อมาจะทำการวัดประสิทธิภาพค่าความสมบูรณ์ของการติดตามและค่าระยะที่ผิดพลาดของการติดตามของแต่ละเส้นทางการเดิน สำหรับการวัดค่าเฉลี่ยค่าความสมบูรณ์ของการติดตามของงานวิจัยที่นำเสนอสำหรับชุดการทดลอง Parking Lot มีค่าเท่ากับ 81.06% ซึ่งค่ามากกว่างานวิจัย GMCP-Tracker[2] ถึง 8.17% สำหรับการทดลอง Town Centre งานวิจัยที่นำเสนอมีค่าเท่ากับ 79.38%

สำหรับเส้นทางการเดินตรงบริเวณที่เกิดการสลับตัวบุคคลพิกัดผลลัพธ์จะมีตำแหน่งที่มีระยะห่างออกไปจากตำแหน่งของคำตอบเส้นทางการเดินที่ถูกต้องของบุคคลที่กำลังติดตามอยู่ จะส่งผลให้การวัดค่าระยะที่ผิดพลาดของการติดตามมีค่าที่มากขึ้นสำหรับเส้นทางการเดินที่มีการสลับตัวบุคคลเกิดขึ้นบ่อยครั้ง เนื่องมาจากค่าระยะที่ผิดพลาดของการติดตามเป็นค่าเฉลี่ยของระยะห่างที่ผิดพลาดไปจากตำแหน่งที่ถูกต้อง ยังมีค่าน้อยจะถือว่ายังมีประสิทธิภาพ สำหรับชุดการทดลอง Parking Lot งานวิจัยที่นำเสนอมีค่าระยะที่ผิดพลาดของการติดตามเท่ากับ 16.17 ซึ่งมีค่าที่ดีกว่างานวิจัย GMCP-Tracker[2] ที่มีค่าเท่ากับ 39.09 สำหรับการทดลอง Town Centre งานวิจัยที่นำเสนอมีค่าเท่ากับ 82.80 ซึ่งมีค่าที่ดีกว่างานวิจัย GMCP-Tracker[2] ที่มีค่าเท่ากับ 129.75 จากการวัดประสิทธิภาพในแง่มุมต่างๆ พบว่างานวิจัยที่นำเสนอสามารถลดจำนวนการติดตามที่เกิดการสลับบุคคลลงได้ และมีตำแหน่งผลลัพธ์ที่ถูกต้องมากยิ่งขึ้น

บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอ การนำผลลัพธ์ของการตรวจจับคนมาทำการเชื่อมโยงข้อมูลเพื่อหาเส้นทางการเดินในระยะสั้นๆ ก่อน ซึ่งในงานวิจัยนี้เรียกว่าแทร็คเล็ต (Tracklet) โดยแต่ละแทร็คเล็ตนั้นจะเป็นผลลัพธ์เส้นทางการเดินช่วงสั้นๆที่ได้จากการเชื่อมโยงข้อมูลภายในช่วงเฟรม โดยช่วงเฟรมที่แบ่งจะแบ่งจำนวนเฟรมเป็นช่วงสั้นๆ ภายใต้สมมุติฐานที่ว่า เป็นช่วงที่คนเดินด้วยความเร็วคงที่และเดินเป็นวิถีเส้นตรง ต่อมางานวิจัยที่นำเสนอจะปรับปรุงประสิทธิภาพแทร็คเล็ตด้วยการประมาณตำแหน่งจากการวิเคราะห์การถดถอยเชิงเส้น แล้วจึงนำแทร็คเล็ตไปรวมเป็นเส้นทางการเดิน (Trajectory) ที่สมบูรณ์ตลอดทั้งวิดีโอ ในขั้นตอนของการเชื่อมโยงข้อมูลเพื่อรวมแทร็คเล็ตเป็นผลลัพธ์เส้นทางการเดิน งานวิจัยที่นำเสนอใช้วิธีการปรับเรียบข้อมูลมาปรับปรุงประสิทธิภาพของการติดตามบุคคลและเพื่อลดปัญหาการติดตามระบุกลับตัวบุคคล

6.1 การวิเคราะห์และสรุปผลการดำเนินงานวิจัย

6.1.1 การวิเคราะห์และสรุปผลค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมสำหรับการจัดค่าที่ไม่มากที่สุด เพื่อเพิ่มประสิทธิภาพการตรวจจับบุคคลจากภาพนิ่ง

การทดลองนี้เป็นการวิเคราะห์พารามิเตอร์ค่าขีดแบ่งพื้นที่ทับซ้อนที่ใช้ในการจัดค่าที่ไม่มากที่สุด เพื่อศึกษาว่าการกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนมีผลต่อแนวโน้มค่าผลบวกจริง (TP) ผลบวกหลง (FP) ผลลบหลง (FN) อย่างไรและศึกษาการเลือกกำหนดค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสม โดยพิจารณาจากค่าความถูกต้องในการตรวจจับหลายวัตถุ (N-MODA) และค่าความแม่นยำในการตรวจจับหลายวัตถุ (N-MODP) เพื่อให้ได้ประสิทธิภาพการตรวจจับที่ดีที่สุด โดยเกณฑ์ที่จะใช้เลือกค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมที่สุด งานวิจัยที่นำเสนอใช้ค่าเฉลี่ยเรขาคณิต (Geometric Mean) ที่พิจารณาจากค่าความถูกต้องในการตรวจจับหลายวัตถุ และค่าความแม่นยำในการตรวจจับหลายวัตถุไปพร้อมกัน เพื่อเป็นเกณฑ์ในการพิจารณาค่าขีดแบ่งพื้นที่ทับซ้อนที่เหมาะสมที่สุด เป็นการสร้างสมดุลระหว่างค่าทั้งสอง จากผลการทดลองพบว่าค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.3 เป็นค่าที่ทำให้ค่าเฉลี่ยเรขาคณิตมีค่ามากที่สุด ดังนั้นค่าขีดแบ่งพื้นที่ทับซ้อนเท่ากับ 0.3 จะเป็นค่าที่เหมาะสมที่สุดในการจัดค่าที่ไม่มากที่สุด ส าหรับการทดลองครั้งนี้

6.1.2 การวิเคราะห์และสรุปผลค่าขีดแบ่งสำหรับการหาค่าที่อยู่ในเกณฑ์และช่วงขอบเขตข้อมูลที่ใช้ในการประมาณเส้นทางการเดิน

การทดลองนี้เป็นการทดลองเพื่อวิเคราะห์หาพารามิเตอร์ที่เหมาะสมของค่าขอบเขตช่วงข้อมูลที่ใช้การประมาณเส้นทางการเดิน และค่าขีดแบ่งระยะห่างในการเลือกพิกัดที่อยู่ในเกณฑ์ โดยเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่บนสื่อออนไลน์ใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้ทดลอง ทาการปรับค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินในช่วง 0.1 ถึง 0.9 ทา การปรับค่าขีดแบ่งระยะห่างในการเลือกค่าที่อยู่ในเกณฑ์เท่ากับ 5 10 และ 15 เพื่อศึกษาว่าค่าช่วงใด ที่เป็นค่าที่เหมาะสม จากผลการทดลอง พบว่าค่าความสมบูรณ์ของการติดตามที่ดีที่สุดเท่ากับ 81.06% ซึ่งเป็นผลลัพธ์ที่ได้จากการกหนดค่าขีดแบ่งระยะห่างในการเลือกค่าที่อยู่ในเกณฑ์เท่ากับ 5 และค่าขอบเขตช่วงข้อมูลที่ใช้ในการประมาณเส้นทางการเดินเท่ากับ 0.2

6.1.3 การวิเคราะห์และสรุปการวัดประสิทธิภาพของการติดตามวัตถุเทียบกับวิธีการ GMCP-Tracker

การวัดประสิทธิภาพของการติดตามวัตถุในงานวิจัยนี้ เพื่อแสดงให้เห็นผลลัพธ์จากงานวิจัยที่นำเสนอที่สามารถช่วยปรับปรุงเส้นทางการเดินให้มีผลลัพธ์ที่มีประสิทธิภาพที่ดีขึ้นและได้ลดปัญหาการติดตามคนเดินสลับบุคคล สำหรับการวัดประสิทธิภาพแบบพิจารณาจากเส้นทางการเดินทั้งหมดตลอดทั้งวิดีโอ การวัดค่าเฉลี่ยค่าความสมบูรณ์ของการติดตามของงานวิจัยที่นำเสนอมีค่าเท่ากับ 81.06% ซึ่งค่ามากกว่างานวิจัย GMCP-Tracker[2] ถึง 8.17% ค่าระยะที่ผิดพลาดของการติดตามเป็นค่าเฉลี่ยของระยะห่างที่ผิดพลาดไปจากตำแหน่งที่ถูกต้อง ยังมีค่าน้อยจะถือว่ายังมีประสิทธิภาพ สำหรับชุดการทดลอง Parking Lot งานวิจัยที่นำเสนอมีค่าระยะที่ผิดพลาดของการติดตามเท่ากับ 16.17 ซึ่งมีค่าที่ดีกว่างานวิจัย GMCP-Tracker[2] ที่มีค่าเท่ากับ 39.09 สำหรับชุดการทดลอง Town Centre งานวิจัยที่นำเสนอมีค่าระยะที่ผิดพลาดของการติดตามเท่ากับ 82.80 ซึ่งมีค่าที่ดีกว่างานวิจัย GMCP-Tracker[2] ที่มีค่าเท่ากับ 129.75 จากการวัดประสิทธิภาพในแง่มุมต่างๆ พบว่างานวิจัยที่นำเสนอสามารถลดจำนวนการติดตามที่เกิดการสลับบุคคลลงได้ และมีตำแหน่งผลลัพธ์ที่ถูกต้องมากยิ่งขึ้น

6.2 ข้อจำกัดและขอบเขตของงานวิจัย

งานวิจัยนี้ได้ทดลองและทาการศึกษาด้วยการติดตามรูปคน ดังนั้นผลลัพธ์ของการติดตามจะเป็นเส้นทางการเดินของบุคคลแต่ละคน งานวิจัยนี้ทาการแบ่งขั้นตอนการหาผลลัพธ์เส้นทางการเดินช่วงสั้นๆ จากช่วงส่วนย่อยวิดีโอก่อน โดยเรียกผลลัพธ์เส้นทางการเดินช่วงสั้นๆนี้ว่าแทร็คเล็ต (Tracklet) โดยแทร็คเล็ตของงานวิจัยนี้มีความหมายว่าเป็นผลลัพธ์เส้นทางการเดินระยะสั้นๆที่ได้จากการเชื่อมโยงข้อมูลวิดีโอภายในช่วงเฟรมระยะสั้นๆที่คนเดินด้วยความเร็วคงที่และเดินเป็นวิถีเส้นตรง หลังจากนั้นจึงค่อยทาการเชื่อมโยงข้อมูลในการรวมแทร็คเล็ตนี้ให้กลายเป็นผลลัพธ์เส้นทางการเดินตลอดทั้งวิดีโอ

6.3 แนวทางในการพัฒนา

เนื่องจากงานวิจัยนี้ ในขั้นตอนการสร้างแทร็คเล็ตจะเชื่อมโยงข้อมูลภายใต้สมมุติฐานที่ว่า เป็นช่วงระยะที่คนเดินด้วยความเร็วคงที่และเดินเป็นวิถีเส้นตรง ซึ่งอาจทาให้ประสิทธิภาพน้อยลงต่อชุดเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการศึกษาไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิดีโอที่คนเดินมีลักษณะเดินเป็นวิถีโค้งมากๆ หรือชุดวิดีโอทดสอบที่มีอัตราเฟรมที่น้อย ดังนั้นแนวทางในการพัฒนาวิธีการติดตามวัตถุในอนาคต อาจเพิ่มขึ้นขั้นตอนการเรียนรู้ของเครื่องที่เพิ่มประสิทธิภาพในการเชื่อมโยงข้อมูล ซึ่งจะสามารถเพิ่มความสามารถในการติดตามที่มากขึ้นและมีความยืดหยุ่นต่อชุดวิดีโอทดสอบที่หลากหลายอีกด้วย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan. "Object Detection with Discriminatively Trained Part-Based Models." in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no.9, Sept. 2010, pp. 1627-1645.
- [2] Roshan Zamir A., Dehghan A., Shah M. "GMCP-tracker: global multi-object tracking using generalized minimum clique graphs." (eds.) Computer Vision – ECCV 2012, vol.7573, 2012 , pp. 343–356.
- [3] L. Fan et al. "A survey on multiple object tracking algorithm." 2016 IEEE International Conference on Information and Automation (ICIA), 2016, pp. 1855-1862.
- [4] J. Zhong, J. Tan, Y. Li, et al. "Multi-Targets Tracking Based On Bipartite Graph Matching." Cybernetics & Information Technologies, vol.14 , no.5, 2014, pp.78-87.
- [5] K. Shafique and M. Shah, "A noniterative greedy algorithm for multiframe point correspondence," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, Jan. 2005, pp. 51-65.
- [6] G. Shu, A. Dehghan, O. Oreifej, E. Hand and M. Shah, "Part-based multiple-person tracking with partial occlusion handling," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 1815-1821.
- [7] Yin, F., Makris, D., Velastin, S. "Performance evaluation of object tracking algorithms". In: Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). Rio de Janeiro, Brazil, (2007)
- [8] R. Kasturi et al., "Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, Feb. 2009, pp. 319-336.
- [9] Bernardin, K., Stiefelhagen, R. "Evaluating multiple object tracking performance: the CLEAR MOT Metrics," in EURASIP Journal on Image And Video Processing, April 2008, pp. 1–10.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1 , 2005, pp. 886-893.
- [11] P. Felzenszwalb and D. Huttenlocher, "Distance Transforms of Sampled Functions," Technical Report 2004-1963, Cornell Univ. CIS, 2004.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ISSN 1738-0073

Colorful DAEGU

The 18th International Symposium on Advanced Intelligent Systems

ISIS2017

October 11-14, 2017 Daegu, Republic of Korea

Sponsored by **KIIS**

Organized by **대구대학교** **전남대학교** **충청대학교**

Co-Sponsored by **IITP** **GVB** **ALLforLAND** **GMT** **YES OF** **DSIC**

Technical co-sponsored by **misotech** **canITcon** **KUMHO SYSTEM** **WRITECH** **FINEMOTORS** **ONT Robot**

www.isis2017.org

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tracklet Estimation using Node Prediction Technique for Effective Human Tracking System

Chaipitak Pattanakittikhun , Orachat Chitsobhuk

Department of Computer Engineering, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.
chaipitak.pat@gmail.com , orachat.ch@kmitl.ac.th

Abstract. Human tracking is an important task in the field of computer vision. One of the popular methods is the trajectory estimation, which generates multiple low-level tracklets and merges them into trajectories. In this paper, an improvement of the tracklet estimation algorithm is proposed. This proposed algorithm consists of four steps: object detection, node association, outlier elimination, and tracklet estimation. First, the proposed algorithm tries to eliminate the outliers, which may lead to estimation failure. Then, node prediction is performed on extracted inliers to supply the predicted nodes and lines, which can be used in the situation of miss detection nodes or used as a whole tracklet prediction. Finally, tracklet generation is performed either from the integration of the extracted inliers of original tracklets and the results of node prediction. In the experiments, the performance of the proposed system is measured in terms of Track Matching Error (TME) and Euclidean distance. It can be seen that the proposed tracklet estimation offers higher accuracy and can help to improve the estimation performance with lower error rates of TME and Euclidean distance by 7.24% and 8.40%, respectively.

Keywords: Human Tracking, Tracklets, Data Association, Linear Regression.

1 Introduction

Object tracking is a significant issue in the field of computer vision. The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video [1]. One of the popular methods is to generate low-level tracklets and merge them into trajectories. In [2], the tracklets were associated by k-NN directed graph to form short tracks and then the Binary Integer Programming method was performed to merge short tracks into full trajectories. [3] generated tracklets using Markov Random Field (MRF) model and uses the Loopy Belief Propagation (LBP) algorithms to merge tracklets to trajectories while [4] linearly estimated tracklets with assumption that each person moves at a constant velocity and merged straight line tracklets into trajectories. This can greatly reduce the complexity of the algorithm.

However, node prediction results rely significantly on the precision of human detection. The inefficient detection can lead to low tracking performance. Miss detection

may occur due to occlusions or false negative detection and this may result in wrong node association and inaccurate tracklet generation. In [4], hypothesis nodes were proposed as additional solutions that can be chosen to improve the node association of a tracklet. However, integrating hypothesis nodes to the estimated tracklet can directly affect the tracklet estimation. In this paper, the tracklet estimation with predicted nodes and lines are proposed to obtain the optimized node prediction. The organizations of this paper are as followed. Section 2 presents the proposed method while Section 3 details the experimental results. Finally, summarize and conclusion is presented in Section 4.

2 Proposed Algorithm

In this paper, object tracking system using proposed node prediction is presented in Fig. 1. The system consists of four main processes: object detection, node association, outlier elimination and tracklet estimation using node prediction.



Fig. 1. Overview of the proposed system

2.1 Object Detection

Object detection is performed to extract the objects of interest from the scene. Several techniques can be used to detect the objects of interest [6, 7]. Histogram of Gradient (HOG) is one of the widely used techniques for human tracking [6]. HOG based human pattern is searched and then the candidate detected objects are classified as human using Support Vector Machine (SVM). However, HOG based technique cannot handle well when the object changes point of view. Another powerful technique adopted in this paper is object detection system using deformable part models [7], which contains two filters: root filter and part filter. The root filter is performed at the global appearance of the objects while the part filters obtain the appearance of each part of the objects.

2.2 Node Association

In this process, detected objects are then associated to construct direction path or tracklet. Several tracklets can be formed as a trajectory of each person along the whole sequence of the video. Graph-based search is adopted to perform node association and estimate the initial tracklet of each object [4]. Linear motion in a short sequence of frames are assumed in order that linear prediction algorithm can be a possible solution. The whole temporal span of a single object in the sequence is incorporated into the data association problem instead of addressing all of them simultaneously to avoid complexity in optimization process [4]. Even though the technique

deals with a single object, all the other objects are implicitly incorporated in graph relation. This makes the algorithm significantly less restrictive.

2.3 Outlier Elimination

Usually, detected objects are very noisy. To enhance the object tracking performance, outliers must be eliminated and only inliers should be extracted and used for tracklet estimation. Object detection usually specifies the location of the detected object at the centroid of the object bounding box. The specified location can be erroneous due to inaccurate detection such as over-detection and under detection. The inaccurate object position can result in great effects on the tracklet estimation. Fig. 2 shows the object location in x-direction (a) and y-direction (b) along each video frame and (c) inlier plot on x-y axis. It can be seen that detection error can lead to outlier track or ineffective tracklet. Therefore, outlier elimination is necessary to clean the data before prediction process is performed. In [4], RANSAC is implemented to extract outliers. However, the algorithm performs the extraction in x-direction and y-direction separately. Due to correlation of object position in x-y direction, performing outlier elimination separately in each direction can lead to incorrect prediction as shown in Fig. 2 (c), where inliers are contaminate with remaining outliers.

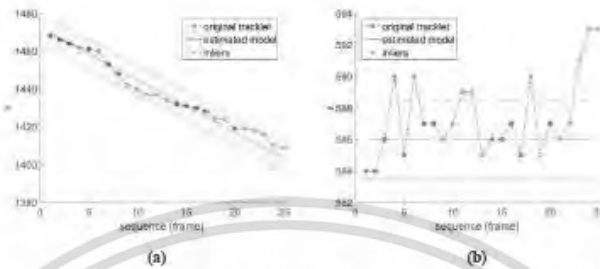
In this paper, linear regression in x-y directions is performed to obtain inlier prediction line. Nodes with the distance greater than τ from the predicted line are considered as outliers and will be discarded. Inliers are the remaining nodes within τ bound. From assumption of linear motion, a tracklet is qualified as a straight line in x-direction and y-direction, (1) and (2), respectively, representing a person moving at a constant velocity. In each frame, we extract inliers and model them using linear least square regression.

$$P_x^i = \hat{w}_{x1}i + \hat{w}_{x0} \quad (1)$$

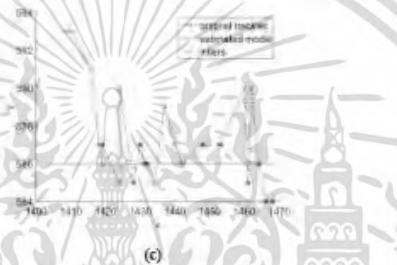
$$P_y^i = \hat{w}_{y1}i + \hat{w}_{y0} \quad (2)$$

where P_x^i is the position of the predicted node in X direction at i^{th} frame, P_y^i is the position of the predicted node in Y direction at i^{th} frame, \hat{w}_{x1} , \hat{w}_{x0} are regression coefficients of the predicted line in X axis, and \hat{w}_{y1} , \hat{w}_{y0} are regression coefficient of the predicted line in Y axis.

The predicted line in X-Y direction in Fig. 4 (c) is considered as 2D motion model of each object in a sequence of a video. However, the nodes along the predicted line must then be related to the time sequence according to constant velocity assumption. The 2D motion model is then separately estimated on X and Y directions, Fig. 4 (a), (b), respectively.



(a) (b)



(c)
 Fig. 2. (a) The inliers of the tracklet in the X direction
 (b) The inliers from tracklet in the Y direction
 (c) The inliers of tracklet in the X-Y direction

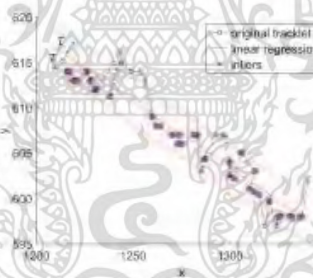


Fig. 3. Finding the inliers of tracklet using the linear least square regression in the x-y direction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

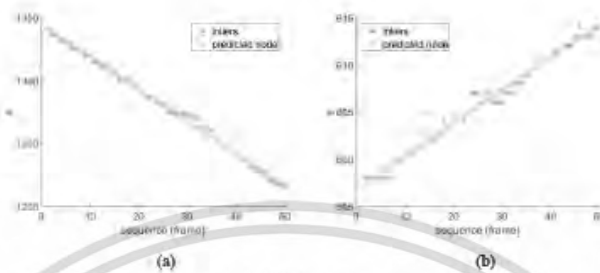


Fig. 4. (a) predicted nodes and estimated tracklet on X-direction
 (b) predicted nodes and estimated tracklet on Y-direction
 (c) predicted node and tracklet on X-Y direction

2.4 Tracklet Estimation

Even though tracklets are usually constructed from actual node association, the detection errors due to occlusion and false negative detection can lead to incorrect tracklet estimation. To enhance the accuracy of the tracklet estimation, the predicted nodes and line are adopted to obtain the optimized node prediction. One solution could be eliminating the outliers and replacing them with the predicted ones at the particular frames. Substitute all predicted nodes on the predicted line to the tracklet could be alternative preference for higher accuracy.

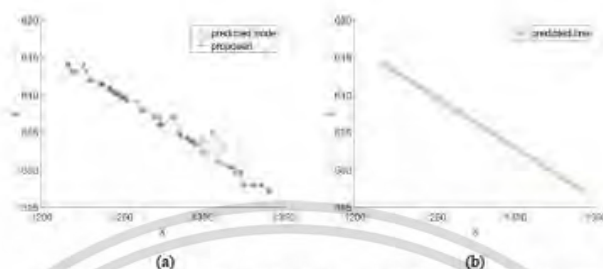


Fig. 5. (a) The tracklet improvement with predicted nodes (b) Predicted Line Preference

3 Experimental Results

In the experiments, the data set of the Parking Lot video is used to evaluate the performance of the human tracking system using proposed node prediction compared to [4]. There are 1000 frames with 1920x1080 pixel resolution. The video sequence is segmented into shots of 50 frames each to ensure linear motion assumption.



Fig. 6. Parking Lot Dataset (a) frame 25 (b) frame 589

Performance measurement is conducted in terms of Track Matching Error (TME) [5] and Euclidean distance as shown in (3) and (4) respectively. TME is used to measure the average distance error per node while Euclidean distance is measured the total errors between the estimated and the ground truth tracklets. The smaller of the TME and the Euclidean distance, the better the accuracy of the tracklet estimation.

$$TME = \frac{\sum_{k=1}^N \sqrt{(X_{GT_{jk}} - X_{ST_{jk}})^2 + (Y_{GT_{jk}} - Y_{ST_{jk}})^2}}{N} \quad (3)$$

$$Euclidean\ Distance = \sqrt{\sum_{k=1}^N (X_{GT_{jk}} - X_{ST_{jk}})^2 + (Y_{GT_{jk}} - Y_{ST_{jk}})^2} \quad (4)$$

where $X_{GT_{ik}}, X_{ST_{jk}}$ are the centroids of the detected objects of the i^{th} ground truth and the j^{th} system tracklet at k^{th} frame in x direction, $Y_{GT_{ik}}, Y_{ST_{jk}}$ are the centroids of the detected objects of the i^{th} ground truth and the j^{th} system tracklet at k^{th} frame in y direction, and N is the number of frames that the i^{th} ground truth intersects with the j^{th} system tracklet.

Table 1. Performance Evaluation of the human tracking systems

Method	TME	Euclidean Distance
Graph-based Human Tracking [4]	591.73	4514.45
Proposed Human Tracking with Predicted Nodes	577.03	4392.57
Proposed Human Tracking with Predicted Line	551.75	4164.35

The performance of the proposed human tracking system with predicted nodes and line is compared to that of [4] and presented in Table I. The experimental results show that the proposed human tracking with predicted nodes provides lower TME and Euclidean distance by 2.54% and 2.77%, respectively, compared to that of [4]. In addition, the proposed human tracking with predicted line substitution provides 7.24% and 8.40% improvement compared to that of [4] in terms of TME and Euclidean distance respectively.

4 Conclusion

This paper proposes tracklet estimation using node prediction technique for effective human tracking system. Outliers are successfully removed according to the assumption of linear motion and a constant velocity. Node prediction is then performed and used as supplement information, which can be integrated to the actual node association. This can help to alleviate the tracking errors due to missed detection and inaccurate node association. From experimental results, it can be seen that the proposed tracklet estimation offers higher accuracy and can help to improve the estimation performance with lower error rates of TME and Euclidean distance by 7.24% and 8.40%, respectively.

Reference

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A Survey. In: ACM Computing Surveys, vol. 38. (2006)
2. Sowah, N.L., Wu, Q., Meng, F., Bo, W., Ngan, K.N.: Strongly connected component multi-object tracking. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC), pp. 394-400. (2016)

3. Zhang, S., Wang, J., Wang, Z., Gong, Y., Lin, Y.: Multi-target tracking by learning local-to-global trajectory models. In: *Pattern Recognition*. vol. 48, pp. 580–590. (2015)
4. Zamir, A.R., Dehghan, A., Shah, M.: GMCP-tracker: global multi-object tracking using generalized minimum clique graphs. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *Computer Vision – ECCV 2012*. LNCS, vol. 7573, pp. 343–356. Springer, Heidelberg (2012)
5. Yin, F., Makris, D., Velastin, S.: Performance evaluation of object tracking algorithms. In: *Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*. Rio de Janeiro, Brazil (2007)
6. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. vol. 1, pp.886–893. San Diego, CA, USA (2005)
7. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object Detection with Discriminatively Trained Part-Based Models. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 32, pp. 1627–45. (2010)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล นายชัยพิทักษ์ พัฒนภิตติคุณ

ที่อยู่ 52/34 ม.5 ซ.ภูมิเวท1 ถ.ภูมิเวท ต.ปากเกร็ด อ.ปากเกร็ด จ.นนทบุรี 11120

ประวัติการศึกษา

พ.ศ.2558 จบการศึกษาวិชากรรมบัณฑิต คณะวิศวกรรมศาสตร์ สาขาวิศวกรรมคอมพิวเตอร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2554 จบการศึกษาจากโรงเรียนเทพศิรินทร์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้