

การอ่านเลขสายรถประจำทางจากภาพ



นายพิชชากร วงศ์ดีตะ

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2560

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

VISION-BASED BUS ROUTE NUMBER READER



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

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2017

Copyright of Chulalongkorn University





# # 5770444521 : MAJOR COMPUTER ENGINEERING

KEYWORDS: IMAGE PRO / MACHINE LEAR / CONVOLUTIONAL NE / FASTER R / BUS NUMBER RECOGNITION / BUS PANEL DETECTION / LOW VISION

PITCHAKORN WONGTA: VISION-BASED BUS ROUTE NUMBER READER. ADVISOR: ASST. PROF. THANARAT CHALIDABHONGSE, 99 pp.

One of the major difficulties for low vision people is to commute independently using public transportation because they can blurry see and recognize buses that are approaching the bus stop but reading bus route number is still challenging to them. Many low vision people need care-takers to help them or ask other people about what route number the bus approaching. Without any assistance, they may get lost or get injured by an accident. This has inspired us to develop this research that assists the low vision people to maintain their independence for traveling using the public transportation.

This research proposes a bus route number recognition system. The method uses machine learning, image processing techniques and the GPS information of the bus stops. The proposed method consists of 4 main steps: the bus panel detection, the text localization, the text recognition and post-processing. In the bus panel detection step, machine learning technique faster R-CNN is applied to build a bus panel detector which localizes the bus panel regions. Text localization based on faster R-CNN is applied to localize text areas. The Google Cloud Vision text recognition and text recognition based on LSTM are used to recognize text after text localization step. Finally, the post-processing is performed to cut the irrelevant characters as well as incorporate GPS information of the image and bus stop to refine the results.

From our experiments, our bus route number recognition yields 62 percent of accuracy. This is from the validation by considering the number of correct bus route number recognition results against the ground truth.

Department: Computer Engineering Student's Signature .....

Field of Study: Computer Engineering Advisor's Signature .....

Academic Year: 2017

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เกิดจากแรงบันดาลใจที่จะช่วยให้ผู้พิการสายตาเลือนรางสามารถเดินทางโดยรถโดยสารประจำทางได้อย่างถูกต้อง

ขอขอบพระคุณ ผศ.ดร.ธนรัตน์ ชลิตาพงศ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยชี้แนะแนวทางในการดำเนินงานวิจัย ให้ข้อคิดต่างๆ ในการทำวิจัย และยังเป็นกำลังใจที่ดีเยี่ยมให้แก่ลูกศิษย์ที่ปรึกษาทุกคน

ขอขอบคุณอาจารย์ และ เจ้าหน้าที่จากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่คอยสนับสนุนการทำงาน ให้กำลังใจในการทำงาน ตลอดจนเพื่อนๆ พี่ๆ น้องๆ ที่ได้ช่วยกันเก็บข้อมูลภาพถ่ายของรถประจำทาง และเป็นกำลังใจอย่างดีเยี่ยมเสมอมา

ขอขอบคุณข้อมูลป้ายรถประจำทาง และสายรถประจำทางในกรุงเทพมหานครและพื้นที่ใกล้เคียงจากแผนที่ลองดู (Longdo Maps) และโครงการ Yak โดย SmartVC ที่ได้ลงพื้นที่ในการเก็บข้อมูลป้ายรถประจำทางและสายรถประจำทางเอง

สุดท้ายขอขอบพระคุณ บิดา มารดา ที่คอยสนับสนุน อบรมสั่งสอนข้าพเจ้ามาโดยตลอด และยังเป็นกำลังใจที่สำคัญที่สุดในการเรียนในระดับปริญญาโทเสมอมา

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญรูปภาพ.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	5
1.3 ขอบเขตของงานวิจัย.....	5
1.4 วิธีการดำเนินงานวิจัย.....	6
1.5 ประโยชน์ที่จะได้รับ.....	6
1.6 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์.....	6
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	8
2.1 ทฤษฎีที่เกี่ยวข้อง.....	8
2.1.1 นิวรอลเน็ตเวิร์ค และ คอนโวลูชันนอลนิวรอลเน็ตเวิร์ค (Neural Network and Convolutional Neural Networks) [4].....	8
2.1.2 การตรวจจับวัตถุโดยวิธีการเสนอบริเวณที่น่าจะเป็นวัตถุ (Object Detection Proposal based Detector) [6].....	12
2.1.2.1 อาร์ชีเอ็นเอ็น (Regions with Convolutional Neural Networks: R-CNN) [8].....	12
2.1.2.2 ฟาส อาร์ชีเอ็นเอ็น (Fast R-CNN) [10].....	13
2.1.2.3 ฟาสเตอร์ อาร์ชีเอ็นเอ็น (Faster R-CNN) [11].....	14

2.1.3 หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory: LSTM) [17].....	20
2.1.4 สูตรฮาเวอ์ไซน์ (Haversine Formula) [19] .....	22
2.1.5 เทนเซอร์โฟล (TensorFlow) [21] .....	23
2.1.6 ระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด (Minimum Editing Distance) [22].....	23
2.1.7 การวัดประสิทธิภาพ (Performance Evaluation).....	26
2.1.7.1 ค่าไอโอยู (Intersection over Union: IoU) [24] .....	26
2.1.7.2 คอนฟิวชันเมทริกซ์ (Confusion Matrix) .....	27
2.1.7.3 ตัววัดประสิทธิภาพการจำแนกโดยคลาส [26] .....	28
2.2 งานวิจัยที่เกี่ยวข้อง.....	29
2.2.1 การตรวจหาเลขสายรถประจำทางโดยใช้เครื่องรับรู้.....	29
2.2.2 การตรวจหาเลขสายรถประจำทางโดยใช้วิธีการประมวลผลภาพ .....	31
2.3 ประเภทของข้อมูลที่ใช้ในงานวิจัย .....	36
2.3.1 บริเวณแผงด้านบนของรถประจำทาง.....	37
2.3.2 ประเภทของแผงด้านบนรถประจำทางที่แสดงข้อมูล.....	37
2.3.3 ประเภทของข้อมูลที่แสดงบนบริเวณแผงด้านบนรถประจำทาง.....	38
2.3.4 ประเภทการแสดงผลข้อมูลบริเวณแผงด้านบนรถประจำทาง .....	39
บทที่ 3 ขั้นตอนและวิธีที่นำเสนอ .....	41
3.1 ขั้นตอนการตรวจจับบริเวณแผงด้านบนของรถประจำทาง .....	43
3.1.1 ขั้นตอนการสอน .....	44
3.1.1.1 การเตรียมชุดข้อมูลสอน.....	44
3.1.1.2 การสร้างตัวตรวจจับแผงด้านบนของรถประจำทาง .....	45
3.1.2 ขั้นตอนการทดสอบ .....	46



3.2 การตรวจหาบริเวณที่มีข้อความ .....	47
3.2.1 กระบวนการตรวจหาบริเวณที่มีข้อความ .....	48
3.3 การรู้จำข้อความ.....	49
3.3.1 ตัวรู้จำข้อความ กูเกิลคลาวด์วิชั่น (Google Cloud Vision Text Recognition) [43].....	50
3.3.2 ตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว [39], [40].....	51
3.4 การประมวลผลหลัง เพื่อปรับปรุงความแม่นยำ.....	53
3.4.1 การตัดตัวอักษรที่ไม่เกี่ยวข้อง .....	54
3.4.2 ข้อมูลจีพีเอสป้ายรถประจำทาง .....	56
บทที่ 4 การทดลองและผลการทดลอง.....	59
4.1 ข้อมูลที่ใช้ในการทดลอง.....	59
4.2 อุปกรณ์ที่ใช้ในการทดลอง.....	60
4.3 การทดลองและผลการทดลอง.....	60
4.3.1 การทดลองตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง .....	62
4.3.2 การทดลองตัวตรวจหาบริเวณที่มีข้อความ .....	66
4.3.3 การทดลองตัวรู้จำข้อความ .....	70
4.3.4 การทดลองหลังจากการทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ .....	72
4.3.5 รวมผลการทดลองทั้งโปรแกรม.....	74
4.4 สรุปผลการทดลอง .....	74
4.4.1 การทดลองขั้นตอนการหาบริเวณแผงด้านบนของรถประจำทาง .....	75
4.4.2 การทดลองขั้นตอนการหาบริเวณที่มีข้อความ .....	75
4.4.3 การทดลองขั้นตอนการรู้จำข้อความ .....	75
4.4.4 การทดลองขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ .....	75

4.4.5 รวมผลการทดลองทั้งโปรแกรม.....	76
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	77
5.1 สรุปผลการวิจัย.....	77
5.2 ข้อเสนอแนะ .....	78
5.2.1 ตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง.....	79
5.2.2 ตัวตรวจจับบริเวณที่มีข้อความ .....	79
5.2.3 ตัวรู้จำข้อความ.....	79
5.2.4. การประมวลผลหลังเพื่อปรับปรุงความแม่นยำ.....	79
รายการอ้างอิง.....	80
ภาคผนวก.....	85
ภาคผนวก ก บทความที่เสนอในงานประชุมวิชาการ.....	86
ประวัติผู้เขียนวิทยานิพนธ์.....	99

## สารบัญตาราง

ตารางที่ 2.1	แสดงตัวอย่างการคำนวณวิธีการแก้ไขที่น้อยที่สุด.....	25
ตารางที่ 2.2	แสดงคอนฟิวชันเมทริกซ์การจำแนกบริเวณแฉงด้านบนของรถประจำทาง.....	28
ตารางที่ 4.1	ชนิดและคุณลักษณะของสมาร์ตโฟนที่ใช้ในการถ่ายภาพ โดยผู้ถ่าย 5 คน .....	60
ตารางที่ 4.2	รายละเอียดของคอมพิวเตอร์ ที่ใช้ในการประมวลผล .....	60
ตารางที่ 4.3	ผลลัพธ์ของการทดลองตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง .....	62
ตารางที่ 4.4	การประเมินประสิทธิภาพของตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง.....	63
ตารางที่ 4.5	ผลการทดลองตัวตรวจหาบริเวณที่มีข้อความ .....	67
ตารางที่ 4.6	การประเมินประสิทธิภาพของตัวตรวจหาบริเวณที่มีข้อความ.....	67
ตารางที่ 4.7	ผลลัพธ์คำตอบที่ได้จากตัวรู้จำกุเกลคลาวด์วิชั่น โดยจำแนกคำตอบเป็น 6 กลุ่ม.....	71
ตารางที่ 4.8	ผลลัพธ์คำตอบ จากผลลัพธ์ไม่มีคำตอบของกุเกลคลาวด์วิชั่น เมื่อเข้าสู่กระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว.....	71
ตารางที่ 4.9	ผลลัพธ์คำตอบโดยรวมทั้งหมดของขั้นตอนการรู้จำข้อความ .....	71
ตารางที่ 4.10	ผลลัพธ์คำตอบจากขั้นตอนการประมวลผลหลังในส่วนขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง .....	72
ตารางที่ 4.11	ตารางคอนฟิวชันเมทริกซ์เปรียบเทียบระหว่างผลลัพธ์คำตอบจากขั้นตอนในการตัดตัวอักษรที่ไม่เกี่ยวข้อง และผลลัพธ์คำตอบหลังจากใช้ข้อมูลจีพีเอสของป้ายรถประจำทาง .....	73
ตารางที่ 4.12	ตารางแสดงผลลัพธ์ของการทดลองทั้งโปรแกรม .....	74

## สารบัญรูปภาพ

รูปที่ 1.1 ตัวอย่างการมองเห็นของผู้พิการสายตาลีอนรางซึ่งมีความหลากหลาย เช่น มีลาน สายตาแคบ หรือ มีจุดพรั่มัวในบางส่วนของลานสายตา [2].....	2
รูปที่ 1.2 ตัวอย่างลักษณะของรถประจำทางในกรุงเทพมหานครที่มีหลากหลายรูปแบบ [3].....	3
รูปที่ 1.3 ความซับซ้อนและหลากหลายของภาพถ่ายรถประจำทาง (ก) ผู้คนบนบั้งรถประจำทาง (ข) เลขสายรถประจำทางไม่ชัดเจนและมีเงาสะท้อนจากกระจก (ค) เลขสายรถประจำทาง แสดงผลโดยหลอดไฟแอลอีดี และมีเงาสะท้อนจากกระจก (ง) เงาสะท้อนจากกระจก (จ) ตัวหนังสือกลืนกับฉากหลัง (ฉ) เลขสายรถประจำทางเลอะเลือน.....	4
รูปที่ 2.1 โครงสร้างของนิวมอลเน็ตเวิร์คโดยทั่วไป .....	8
รูปที่ 2.2 โครงสร้างของคอนโวลูชันนอลนิวมอลเน็ตเวิร์ค [5].....	9
รูปที่ 2.3 วิธีการคอนโวลูชัน.....	10
รูปที่ 2.4 ตัวอย่างกระบวนการการสุ่มตัวอย่าง.....	11
รูปที่ 2.5 กระบวนการอาร์ซีเอ็นเอ็น [9].....	13
รูปที่ 2.6 กระบวนการฟาส อาร์ซีเอ็นเอ็น [10].....	14
รูปที่ 2.7 โครงสร้างของฟาสเตอร์อาร์ซีเอ็นเอ็น .....	15
รูปที่ 2.8 ตัวอย่างโครงสร้างของ เรสเน็ต [14] .....	15
รูปที่ 2.9 เปรียบเทียบโครงสร้างคอนโวลูชันนอลนิวมอลเน็ตเวิร์ค ซ้าย: โครงสร้างคอนโวลูชัน นอลนิวมอลเน็ตเวิร์คทั่วไป ขวา: โครงสร้างคอนโวลูชันนอลนิวมอลเน็ตเวิร์คแบบเรสลิดูบล็อค (Residual Block) [16].....	16
รูปที่ 2.10 กระบวนการเครือข่ายบริเวณที่เสนอ .....	17
รูปที่ 2.11 กระบวนการเครือข่ายบริเวณที่เสนอ โดยข้อมูลของแต่ละแองเคิล มีข้อมูลการจำแนก วัตถุที่เสนอ และข้อมูลพิกัดบริเวณที่สนใจ.....	18
รูปที่ 2.12 กระบวนการอาร์โอไอ พูลลิง (ขั้นต้น) .....	18
รูปที่ 2.13 กระบวนการอาร์โอไอ พูลลิง (ขั้นหลัง) .....	19
รูปที่ 2.14 กระบวนการจำแนก (ผลลัพธ์ที่เป็นค่าของหมวดหมู่ของวัตถุ).....	19

รูปที่ 2.15	กระบวนการจำแนก (แก้ไขพิกัดให้ใกล้เคียงกับบริเวณที่มีวัตถุ).....	20
รูปที่ 2.16	การทำงานของหน่วยความจำระยะสั้นระยะยาว [18].....	21
รูปที่ 2.17	ตัวอย่างการแก้ไขด้วยจำนวนน้อยที่สุด (ก) ตัวอย่างสตรีง 2 ชุดที่ต้องการแก้ไข โดยจะแก้ไขคำจากทางซ้าย ไปเป็นคำทางด้านขวา (ข) กระบวนการทำการแก้ไขจากตัวอย่างสตรีงในรูป (ก).....	24
รูปที่ 2.18	แสดงถึงโปรแกรมการคำนวณการแก้ไขด้วยจำนวนน้อยที่สุด .....	25
รูปที่ 2.19	กระบวนการจับคู่กันของลำดับ [23].....	26
รูปที่ 2.20	แสดงตัวอย่างของการวัดผลค่าไอโอยู .....	27
รูปที่ 2.21	การทำงานของระบบตรวจจับรถประจำทางโดยใช้เครื่องรับรู้อาร์เอฟไอดี [28] .....	30
รูปที่ 2.22	ขั้นตอนการทำงานของระบบตรวจจับรถประจำทางและการอ่านเลขสายรถประจำทาง [29].....	31
รูปที่ 2.23	รูปนำเข้าจากกล้องที่ติดบริเวณป้ายรถประจำทาง [31] .....	32
รูปที่ 2.24	การตรวจจับรถประจำทาง (a) ภาพลำดับก่อนหน้า (b) ภาพลำดับปัจจุบัน (c) ภาพความต่างของ (a)(b), (d) ภาพไบนารีโดยจากค่าขีดแบ่ง [32].....	33
รูปที่ 2.25	ภาพนำเข้าเป็นภาพจากกล้อง 2 ตัวที่เชื่อมกัน [33].....	34
รูปที่ 2.26	ระบบตรวจจับเลขสายรถประจำทางจากภาพถ่ายวิดีโอ (a) ภาพลำดับนำเข้า (b) ภาพไบนารี [34].....	35
รูปที่ 2.27	ระบบตรวจจับและรู้จำรถประจำทาง (a) ภาพลำดับนำเข้า (b) ภาพจากวิธีการ LSD [35].....	35
รูปที่ 2.28	ระบบตรวจจับและรู้จำเลขสายรถประจำทาง (a) ภาพลำดับนำเข้า (b) เลือกรูปพื้นหน้าบริเวณรถประจำทางเป็นสีแดง (c) ภาพไบนารีหลังจากการทำ Otsu Threshold [36].....	36
รูปที่ 2.29	บริเวณแฉกด้านบนของรถประจำทาง ที่แสดงเฉพาะข้อมูลสายรถประจำทาง .....	37
รูปที่ 2.30	แฉกด้านบนรถประจำทางที่ติดกับตัวรถ .....	37
รูปที่ 2.31	แฉกด้านบนรถประจำทางที่ยื่นออกมาจากตัวรถ .....	38
รูปที่ 2.32	รถประจำทางที่ข้อมูลบริเวณแฉกด้านบนของรถประจำทางปรากฏเพียงตัวเลขสายรถประจำทาง .....	38

รูปที่ 2.33	รถประจำทางที่ข้อมูลบริเวณแผงด้านบนของรถประจำทางมีเลขสายรถประจำทาง และข้อมูลสถานที่ที่ผ่าน .....	39
รูปที่ 2.34	รถประจำทางที่ข้อมูลบริเวณแผงด้านบนของรถประจำทางมีเลขสายรถประจำทางกับ ตัวอักษรที่มีความหมายพิเศษ .....	39
รูปที่ 2.35	รถประจำทางที่แสดงข้อมูลบริเวณแผงด้านบนของรถประจำทางแบบ พิมพ์หรือทาสี ถาวร .....	40
รูปที่ 2.36	รถประจำทางที่แสดงข้อมูลบริเวณแผงด้านบนของรถประจำทางแบบ หลอดไฟ แอลอีดี .....	40
รูปที่ 3.1	แผนผังการทำงานของระบบการอ่านเลขสายรถประจำทางจากภาพ .....	41
รูปที่ 3.2	ตัวอย่างภาพถ่ายรถประจำทาง ชสมก. หรือรถร่วมบริการ ที่ใช้ในวิทยานิพนธ์ฉบับนี้ ซึ่ง มีความหลากหลายและซับซ้อนของทั้งฉากหลัง รวมถึงการบังจากอ็อบเจกต์ต่าง ๆ (ก)(ข)(ค) (ง) แสดงภาพถ่ายรถประจำทางที่ในภาพประกอบไปด้วยสิ่งทีนอกเหนือจากรถประจำทาง เช่น รถยนต์ บุคคล หรือป้ายโฆษณา .....	42
รูปที่ 3.3	ขั้นตอนของการตรวจจับบริเวณแผงด้านบนของรถประจำทาง .....	43
รูปที่ 3.4	โปรแกรม labelling ที่ใช้ในการเก็บพิกัดข้อมูล .....	44
รูปที่ 3.5	กราฟค่าความเสียหายทั้งหมด ของการสอน .....	45
รูปที่ 3.6	การทดสอบตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง (ก) ขั้นตอนการทดสอบ (ข) ตัวอย่างรูปภาพหลังจากเข้าตัวตรวจจับแผงด้านบนของรถประจำทาง โดยกรอบสี่เหลี่ยมสีเขียวแสดงถึงบริเวณแผงด้านบนของรถประจำทาง .....	46
รูปที่ 3.7	รูปที่ได้หลังจากการตรวจจับบริเวณแผงด้านบนของรถประจำทาง .....	47
รูปที่ 3.8	แสดงขั้นตอนการทำงานของตรวจหาบริเวณที่มีข้อความ .....	47
รูปที่ 3.9	ตัวอย่างผลลัพธ์ที่ได้จากการตรวจหาบริเวณที่มีข้อความ (ก)(ข) ตัวอย่างที่ได้จากการ ตรวจหาบริเวณที่มีข้อความที่บริเวณแผงด้านบนรถประจำทางแสดงเฉพาะเลขสายรถประจำทาง (ค) ตัวอย่างที่ได้จากการตรวจหาบริเวณที่มีข้อความที่บริเวณแผงด้านบนรถประจำทางแสดงเลข สายรถประจำทาง และชื่อสถานที่ .....	48
รูปที่ 3.10	ขั้นตอนการทำงานของรู้จำข้อความ .....	49
รูปที่ 3.11	ขั้นตอนการทำงานของตัวรู้จำข้อความจากกุเกิลคลาวด์วิชั่น .....	50

รูปที่ 3.12 ตัวอย่างคำตอบที่ได้จากภูเกิลคลาวด์วิชั่น .....	51
รูปที่ 3.13 ขั้นตอนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว.....	52
รูปที่ 3.14 ลำดับขั้นตอนของการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ .....	53
รูปที่ 3.15 ลักษณะรูปแบบตัวหนังสือบริเวณแฉกด้านบนของรถประจำทาง (ก) แบบตัวเลขสายรถประจำทางอย่างเดียว (ข) แบบตัวเลขสายรถประจำทางกับข้อมูลสถานที่ที่ผ่าน (ค) แบบตัวเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ .....	55
รูปที่ 3.16 ข้อมูลป้ายรถประจำทางทั่วกรุงเทพมหานครและปริมณฑล จากแผนที่ลองดู [47].....	57
รูปที่ 3.17 แผนที่ลองดูที่แสดงข้อมูลป้ายรถประจำทาง ที่มีข้อมูลของสายรถประจำทางที่ผ่าน [47].....	57
รูปที่ 4.1 บริเวณจุดที่ทำการถ่ายภาพรถประจำทาง บริเวณป้ายรถประจำทาง .....	59
รูปที่ 4.2 ตัวอย่างภาพถ่ายที่ได้ทำการรวบรวม ที่มีลักษณะของภาพที่หลากหลาย (ก) ภาพถ่ายรถประจำทางที่มีผู้คนบดบังรถประจำทาง (ข) ภาพถ่ายรถประจำทางที่มีเงาสะท้อนบริเวณแฉกด้านบนรถประจำทาง (ค) ภาพถ่ายรถประจำทางที่ในภาพมีรถประจำทางมากกว่า 1 คัน .....	61
รูปที่ 4.3 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแฉกด้านบนของรถประจำทาง (ก)-(ง) ผลลัพธ์ที่ถูกต้อง (TP) ที่ได้จากการทดลองของตัวตรวจจับแฉกด้านบนของรถประจำทาง .....	63
รูปที่ 4.4 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแฉกด้านบนของรถประจำทาง (ก)-(ง) ผลลัพธ์ที่ถูกต้อง (TP) ที่ได้จากการทดลองของตัวตรวจจับบริเวณแฉกด้านบนของรถประจำทาง.....	64
รูปที่ 4.5 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแฉกด้านบนของรถประจำทางที่ผิดพลาด สีฟ้า คือบริเวณที่เป็นค่าความจริง, สีเขียว คือบริเวณที่ตรวจจับได้ (ก)(ข) ตัวตรวจจับ ตรวจจับได้นอกเหนือบริเวณที่เป็นแฉกด้านบนรถประจำทาง (ค) ตรวจจับไม่พบบริเวณแฉกด้านบนรถประจำทาง .....	65
รูปที่ 4.6 ผลการทดลองของตัวตรวจหาบริเวณที่มีข้อความ (ก)(ค)(ง) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหาบริเวณที่มีข้อความ ซึ่งแฉกด้านบนรถประจำทางแสดงเลขสายรถประจำทางเพียงอย่างเดียว (ข) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหาบริเวณที่มีข้อความ ซึ่งแฉกด้านบนรถประจำทางแสดงเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ (จ)(ฉ) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหาบริเวณที่มีข้อความ ซึ่งแฉกด้านบนรถประจำทางแสดงเลขสายรถประจำทางและข้อมูลสถานที่ที่ผ่าน .....	66

**รูปที่ 4.7** ตัวอย่างความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีที่แผงด้านบนของรถประจำทางมีข้อความอื่นๆ นอกจากข้อความที่เป็นข้อมูลสายรถประจำทาง หรือ สภาพแวดล้อมสิ่งของที่มีลักษณะที่คล้ายกับตัวอักษรหรือข้อความ (ก)(จ) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความอื่นนอกจากเลขสายรถประจำทาง (ข)(ง) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความจากป้ายโฆษณา (ค) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความที่ควรเป็น 1 บริเวณ แต่ตรวจจับได้ 2 บริเวณ และอีก 1 บริเวณที่ตรวจจับผิด .....

68

**รูปที่ 4.8** ตัวอย่างความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีที่ไม่สามารถตรวจจับตัวอักษรได้ (ก)(ข)(ค)(จ) คือภาพที่มีลักษณะเป็นตัวเลขเพียงตัวเดียวซึ่งตัวตรวจหาบริเวณที่มีข้อความไม่สามารถตรวจจับได้เนื่องจากสมมุติฐานของตัวตรวจจับข้อความ คือข้อความจะประกอบด้วยตัวอักษรหลายตัวเรียงต่อกัน (ง) ตัวตรวจจับข้อความตรวจจับผิดพลาด ..

69





# บทที่ 1

## บทนำ

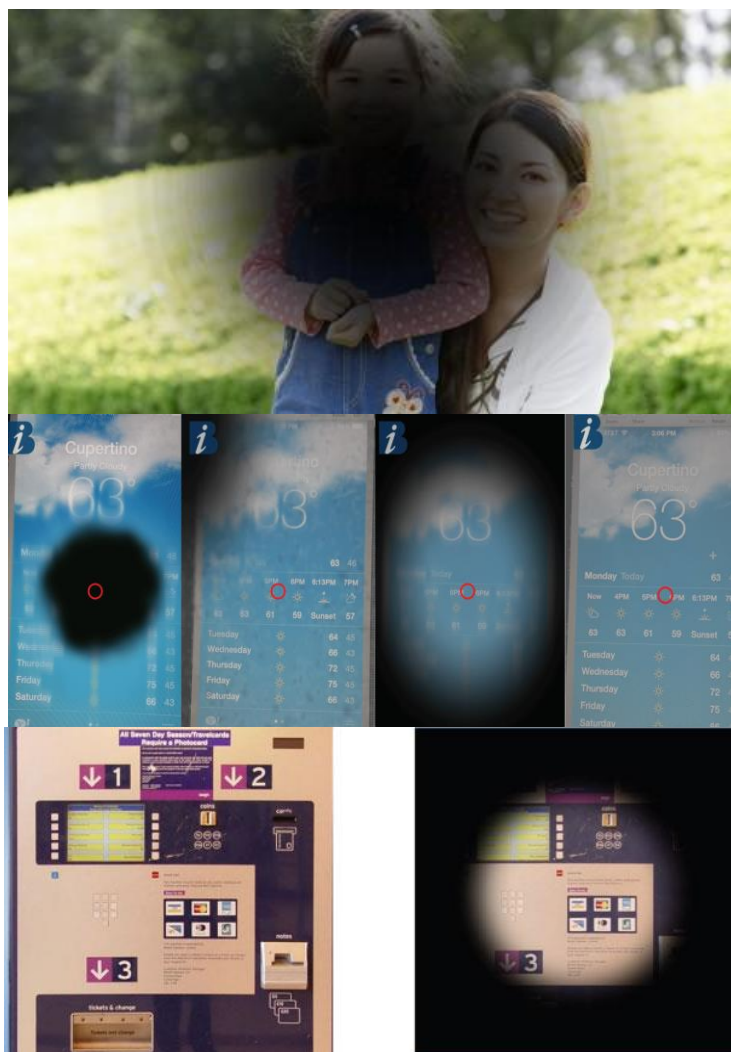
### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในสังคมปัจจุบัน เทคโนโลยีได้เข้ามามีบทบาทกับมนุษย์ในด้านต่าง ๆ มากมาย โดยถูกนำมาใช้เพื่อช่วยเหลือมนุษย์ให้สามารถใช้ชีวิตได้อย่างมีประสิทธิภาพและสะดวกสบายมากยิ่งขึ้น อย่างไรก็ตามก็ยังมีกลุ่มบุคคลอีกกลุ่มหนึ่งคือกลุ่มคนพิการ เช่น ผู้พิการทางสายตา ผู้พิการทางด้านสติปัญญา ผู้พิการทางการได้ยิน ฯลฯ ที่ยังประสบความยากลำบากในการใช้ชีวิตประจำวัน โดยเฉพาะอย่างยิ่งผู้พิการทางสายตาซึ่งไม่สามารถมองเห็นสิ่งต่าง ๆ ได้อย่างชัดเจน รวมถึงผู้พิการทางสายตาที่ตาบอดสนิท กลุ่มคนเหล่านี้มักประสบปัญหาในการเดินทางไปยังสถานที่ต่าง ๆ จนอาจนำไปสู่การได้รับอุบัติเหตุ การนำเทคโนโลยีมาช่วยเหลือผู้พิการทางสายตานั้นจึงเป็นแนวคิดที่น่าสนใจ นำไปสู่การยกระดับคุณภาพชีวิตให้ผู้พิการทางสายตาให้สามารถใช้ชีวิตได้สะดวกยิ่งขึ้น

การจำแนกประเภทโรคและปัญหาสุขภาพที่เกี่ยวข้องระหว่างประเทศครั้งที่ 10 (ICD-10) [1] ขององค์การอนามัยโลก พบว่าโรคทางสายตาสามารถจำแนกออกเป็น 4 หมวด คือ สายตาสั้น สายตาวอก สายตาบอดสนิท และสายตาที่บอดสนิท โดยสายตาเลือนรางจะจัดอยู่ในระหว่างกลุ่มของสายตาบอดสนิทและกลุ่มของสายตาบอดสนิท หมายถึงกลุ่มคนพิการสายตาเลือนรางเหล่านี้สามารถมองเห็นภาพได้บางส่วน อาจมองเห็นภาพที่มีลักษณะพร่ามัวหรือมองได้แคบ สาเหตุของสายตาเลือนรางมักเกิดจากการที่ได้รับบาดเจ็บที่ดวงตาหรือโรคประจำตัวที่อาจส่งผลกระทบต่อดวงตา เช่น โรคเบาหวาน โรคต้อหิน โรคต้อกระจก และจอประสาทตาเสื่อม สายตาเลือนรางสามารถเกิดได้ในทุกเพศ ทุกวัย โดยส่วนมากมักเกิดกับผู้สูงอายุ จากการสำรวจจำนวนผู้พิการสายตาเลือนรางขององค์การอนามัยโลก ในปี 2560 พบว่ามีจำนวนมากถึง 217 ล้านคน โดยในประเทศไทยมีกลุ่มคนที่สายตาเลือนรางประมาณ 300,000 คน จากข้อมูลของสมาคมคนสายตาเลือนรางแห่งประเทศไทยและยังมีจำนวนเพิ่มขึ้นเรื่อย ๆ รูปที่ 1.1 แสดงตัวอย่างการมองเห็นของผู้พิการสายตาเลือนรางที่มองเห็นได้แคบ หรือมีบางสิ่งมาบดบังสายตา

ผู้พิการสายตาเลือนรางนั้นมักประสบปัญหาในการใช้ชีวิตประจำวันต่าง ๆ เช่น การอ่านหนังสือ ดูโทรทัศน์ หรือแม้กระทั่งการเขียน โดยที่พวกเขาต้องใช้เวลาในการเรียนรู้หรือหาวิธีที่ทำให้สามารถทำกิจกรรมเหล่านี้ได้อย่างเป็นปกติ ปัจจุบันได้มีอุปกรณ์ที่จัดทำขึ้นมาเพื่อช่วยเหลือกลุ่มผู้พิการสายตาเลือนรางมากมาย ตัวอย่างเช่น แว่นขยายพกพา แว่นขยายอิเล็กทรอนิกส์ที่สามารถปรับภาพและสีได้ และอุปกรณ์ช่วยอ่านหนังสือ แต่อุปกรณ์ส่วนใหญ่ยังยากต่อการใช้งานและยังมีราคาที่สูง ผู้พิการสายตาเลือนรางส่วนมากจึงจำเป็นต้องมีผู้ดูแลเพื่อช่วยเหลือในการทำกิจวัตรประจำวัน

ต่าง ๆ แต่ยังมีกลุ่มผู้พิการสายตาเลือนรางอีกกลุ่มที่ยังเป็นวัยรุ่นหรือวัยกลางคนที่ต้องการพึ่งตนเอง โดยการใช้ประโยชน์จากเทคโนโลยีที่มีอยู่ในปัจจุบันเข้ามาช่วยในการอำนวยความสะดวก ตัวอย่างเช่น การใช้สมาร์ทโฟนทำการถ่ายภาพและทำการขยายภาพเพื่อช่วยในการอ่าน เป็นต้น ทั้งนี้ ผู้พิการสายตาเลือนรางต้องมีความสามารถในการใช้เทคโนโลยีเหล่านี้ด้วย



รูปที่ 1.1 ตัวอย่างการมองเห็นของผู้พิการสายตาเลือนรางซึ่งมีความหลากหลาย เช่น มีลานสายตาแคบ หรือ มีจุดพรั่มัวในบางส่วนของลานสายตา [2]

การเดินทางไปยังสถานที่ต่าง ๆ ถือเป็นกิจกรรมประจำวันอย่างหนึ่งของผู้คน สามารถทำได้หลายรูปแบบ เช่น การเดิน การวิ่ง การขับรถยนต์ส่วนบุคคล หรือ ใช้บริการขนส่งสาธารณะ เป็นต้น ในกรุงเทพมหานครถือว่าเป็นเมืองที่ถูกพูดถึงว่ารถติดที่สุดเป็นอันดับต้นๆ ของโลก การใช้บริการ

ขนส่งสาธารณะในกรุงเทพมหานครจึงเป็นทางเลือกหนึ่งในการเดินทาง เช่น การเดินทางโดย รถไฟฟ้า หรือรถประจำทาง

รถประจำทางนั้นเป็นการเดินทางที่ง่ายและประหยัด คนทุกระดับสามารถเดินทางโดยรถประจำทางได้ การใช้บริการขนส่งสาธารณะเป็นปัญหาใหญ่อีกปัญหาหนึ่งที่ผู้พิการสายตาเลือนรางประสบในการเดินทางไปยังสถานที่ต่างๆ โดยปกติแล้วผู้พิการสายตาเลือนรางที่ใช้บริการรถประจำทางอาจมีผู้ช่วยประจำตัวหรือต้องสอบถามบุคคลที่ยืนรอรถประจำทางว่ารถประจำทางที่เข้ามาจอดที่ป้ายเป็นสายใด เนื่องจากผู้พิการสายตาเลือนรางไม่สามารถอ่านตัวเลขสายรถประจำทางได้อย่างชัดเจนหรืออ่านผิดพลาด หากอยู่ลำพังอาจทำให้เกิดการขึ้นรถประจำทางผิด ส่งผลให้เกิดการพลัดหลง เสียเวลา หรืออาจส่งผลร้ายแรงถึงชีวิต และทรัพย์สินของตัวผู้พิการสายตาเลือนรางได้ และในยุคที่ทุกวันนี้ทุกคนสามารถถ่ายรูปได้จากสมาร์ทโฟน จึงเป็นแรงบันดาลใจให้ผู้วิจัยได้จัดทำระบบการอ่านเลขสายรถประจำทางจากภาพถ่าย เพื่อช่วยให้ผู้พิการสายตาเลือนรางสามารถขึ้นรถประจำทางเองได้อย่างถูกต้อง ซึ่งเป็นปัญหาที่สำคัญและต้องได้รับการช่วยเหลือ



รูปที่ 1.2 ตัวอย่างลักษณะของรถประจำทางในกรุงเทพมหานครที่มีหลากหลายรูปแบบ [3]



(ก)



(ข)



(ค)



(ง)



(จ)



(ฉ)

รูปที่ 1.3 ความซ้ำซ้อนและหลากหลายของภาพถ่ายรถประจำทาง (ก) ผู้คนบดบังรถประจำทาง (ข) เลขสายรถประจำทางไม่ชัดเจนและมีเงาสะท้อนจากกระจก (ค) เลขสายรถประจำทางแสดงผลโดยหลอดไฟแอลอีดี และมีเงาสะท้อนจากกระจก (ง) เงาสะท้อนจากกระจก (จ) ตัวหนังสือสีกลิ้งกับฉากหลัง (ฉ) เลขสายรถประจำทางเลอะเลือน

สำหรับแนวทางการระบุวารรถประจำทางเป็นสายไคนั้น จากการทบทวนวรรณกรรมพบว่า ในต่างประเทศมีวิธีการใช้เครื่องรับรู้อาร์เอฟไอดี (RFID) ในการระบุสายรถประจำทาง โดยต้องมีการติดตั้งอุปกรณ์พิเศษในการอ่านค่าจากเครื่องรับรู้อาร์เอฟไอดี ร่วมกับระบบฐานข้อมูลของสายรถประจำทาง ซึ่งวิธีดังกล่าวต้องอาศัยงบประมาณมาก เพราะต้องติดตั้งเครื่องรับรู้จำนวนมาก อีกวิธีหนึ่งคือการใช้เทคโนโลยีในการประมวลผลทางภาพในการตรวจจับรถประจำทาง โดยวิธีนี้มีข้อดีคือไม่ต้องใช้งบประมาณมากเพราะไม่จำเป็นต้องติดตั้งเครื่องรับรู้ทุกคัน สามารถใช้เทคโนโลยีประมวลผลภาพและการเรียนรู้ของเครื่องในการตรวจจับสายรถประจำทาง

เนื่องจากรถประจำทางในกรุงเทพมหานครนั้นมีรูปแบบที่หลากหลาย (เช่น รูปที่ 1.2) ทั้งตัวรถประจำทางที่มาจากอู่รถประจำทางต่างๆ ที่มีลักษณะแตกต่างกัน สีของรถประจำทางที่มีมากมายหลายสี ลักษณะตัวหนังสือของรถประจำทางแต่ละคันแตกต่างกัน ตำแหน่งของเลขรถประจำทางที่อยู่คนละตำแหน่ง การแสดงผลของตัวเลขสายรถประจำทางที่มีหลายแบบ เช่น แบบพ่นสีหรือถาวร และแบบแสดงผลโดยหลอดไฟแอลอีดี รวมถึงลักษณะของภาพถ่ายที่มีสิ่งรบกวนอยู่หลายปัจจัย เช่น สภาพของแสง การขาดหายของตัวเลขสายรถประจำทาง การสะท้อนจากกระจกรถประจำทาง มีต้นไม้หรือสิ่งอื่น ๆ ที่บดบังรถประจำทาง เป็นต้น เนื่องด้วยปัจจัยเหล่านี้ การสกัดเลขสายรถประจำทางออกมาได้อย่างแม่นยำจึงมีความท้าทายเป็นอย่างยิ่ง โดยในรูปที่ 1.3 ได้แสดงความหลากหลายของภาพถ่ายรถประจำทางในสถานการณ์ที่แตกต่างกัน

จากปัญหาเหล่านี้ ผู้วิจัยต้องการพัฒนาระเบียบขั้นตอนวิธีในการอ่านหมายเลขรถประจำทางจากภาพถ่าย โดยมุ่งหวังว่า ในอนาคตสามารถนำไปต่อยอดเพื่อพัฒนาระบบต้นแบบบนอุปกรณ์พกพา เช่น โทรศัพท์สมาร์ทโฟน เพื่อให้ผู้พิการสายตาเลือนรางสามารถนำไปใช้ได้ต่อไป

## 1.2 วัตถุประสงค์ของงานวิจัย

เพื่อพัฒนาขั้นตอนวิธีที่สามารถ รู้จำเลขสายรถประจำทางจากภาพถ่าย

## 1.3 ขอบเขตของงานวิจัย

1.3.1 ภาพนำเข้รวบรวมจากภาพถ่าย โดยผู้ถ่ายจำนวนไม่น้อยกว่า 5 คน ด้วยสมาร์ทโฟนของตนเอง และตั้งค่ากล้องตาม Default (เพื่อให้ได้ภาพถ่ายหลากหลายมุมมองและจากหลากหลายแหล่งกำเนิด)

1.3.2 ภาพนำเข้ที่จะใช้ในการทดสอบเป็นภาพถ่ายรถประจำทางของ ขสมก. หรือรถร่วมที่วิ่งในกรุงเทพมหานคร

1.3.3 ภาพนำเข้เป็นภาพถ่าย ในเวลากลางวัน

- 1.3.4 ภาพนำเข้าจะต้องปรากฏเลขสายรถประจำทางที่มนุษย์สามารถมองเห็นได้
- 1.3.5 มุมมองของการถ่ายภาพรถประจำทาง จะต้องเป็นมุมที่ผู้ถ่ายยืนอยู่บริเวณป้ายรถประจำทาง แล้วถ่ายออกไป
- 1.3.6 โปรแกรมต้นแบบทำงานบน Laptop/PC
- 1.3.7 ภาพนำเข้าอาจมีรถประจำทางมากกว่า 1 คันได้
- 1.3.8 จำนวนภาพนำเข้าที่จะใช้ในการทดสอบ จะต้องมีความเพียงพอต่อการทดลองความสามารถของระบบต้นแบบ

#### 1.4 วิธีการดำเนินงานวิจัย

- 1.4.1 รวบรวมภาพถ่ายจากผู้ถ่ายภาพทั้งหมด 5 คน ถ่ายด้วยกล้องสมาร์ทโฟนที่ตั้งค่ากล้องแบบ Default และผู้ถ่ายยืนอยู่บริเวณป้ายรถประจำทางแล้วถ่ายออกไป
- 1.4.2 ออกแบบวิธีสกัดบริเวณแผงด้านบนของรถประจำทาง
- 1.4.3 ทดสอบวิธีสกัดบริเวณแผงด้านบนของรถประจำทาง
- 1.4.4 ทดสอบวิธีการหาบริเวณที่มีข้อความ จากรูปบริเวณแผงด้านบนรถประจำทาง
- 1.4.5 ทดสอบวิธีการรู้จำเลขสายรถประจำทาง
- 1.4.6 ออกแบบวิธีประมวลผลหลัง
- 1.4.7 ทดสอบวิธีประมวลผลหลัง
- 1.4.8 สรุปผลงานวิจัย

#### 1.5 ประโยชน์ที่จะได้รับ

- 1.5.1 งานวิจัยนี้นำเสนอระบบอ่านเลขสายรถประจำทางจากภาพถ่าย ซึ่งระบบสามารถวิเคราะห์หาเลขสายรถประจำทางได้
- 1.5.2 ระบบนี้เป็นต้นแบบในการต่อยอดเพื่อพัฒนาต้นแบบบนอุปกรณ์พกพา เช่น โทรศัพท์สมาร์ทโฟน เพื่อให้ผู้พิการสายตาเลือนรางสามารถนำไปใช้จริงได้ต่อไป

#### 1.6 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

P. Wongta, T. Kobchaisawat and T.H. Chalidabhongse, “An Automatic Bus Route Number Recognition,” in the 13<sup>th</sup> International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016.

P. Wongta and T.H. Chalidabhongse, “Vision-Based Bus Route Number Reader for Visually Impaired Travelers,” in the 2<sup>nd</sup> International Conference on Imaging, Signal Processing and Communication (ICISPC), 2018.



## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

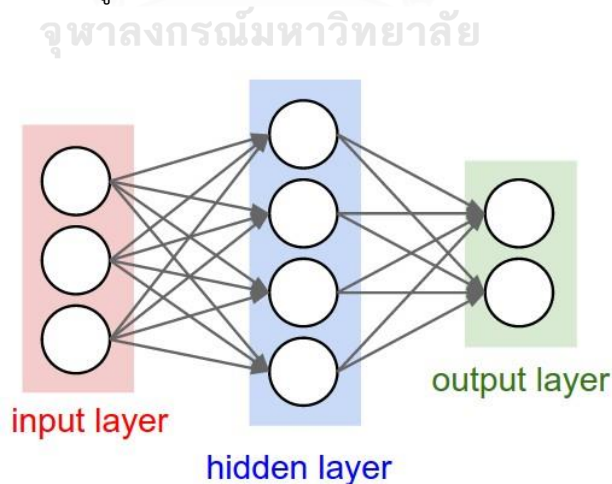
ในส่วนของทฤษฎีและงานวิจัยที่เกี่ยวข้อง กล่าวถึงทฤษฎีที่เกี่ยวข้องที่ใช้เป็นหลักในวิทยานิพนธ์ฉบับนี้ และงานวิจัยที่เกี่ยวข้องกับการตรวจจับเลขสายรถประจำทางที่ใช้ในต่างประเทศ รวมถึงประเภทของรถประจำทางที่ให้บริการในกรุงเทพมหานคร ทางด้านต่างๆ

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 นิวรอลเน็ตเวิร์ค และ คอนโวลูชันนอลนิวรอลเน็ตเวิร์ค (Neural Network and Convolutional Neural Networks) [4]

นิวรอลเน็ตเวิร์ค (Neural Network) คือ ระบบโครงข่ายประสาทเทียม ที่จำลองมาจากสมองของมนุษย์ โดยสมองของมนุษย์จะมีหน่วยการประมวลผลเล็กๆ จำนวนมาก และเชื่อมโยงถึงกันด้วยโครงข่ายระบบประสาท ทำให้มนุษย์สามารถคิดวิเคราะห์ จำแนกข้อมูลได้อย่างรวดเร็ว นิวรอลเน็ตเวิร์คจึงจำลองแนวทางการเรียนรู้ของมนุษย์ และนำไปใช้กับระบบคอมพิวเตอร์ให้มีความสามารถในการเรียนรู้ และคิดวิเคราะห์ได้

นิวรอลเน็ตเวิร์ค มีโครงสร้างประกอบไปด้วย 3 ส่วนคือ ชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) ชั้นผลลัพธ์ (Output Layer) โดยชั้นนำเข้า เป็นข้อมูลของคุณลักษณะสำคัญ (Feature) แบบ 1 มิติ แสดงดังรูปที่ 2.1



รูปที่ 2.1 โครงสร้างของนิวรอลเน็ตเวิร์คโดยทั่วไป

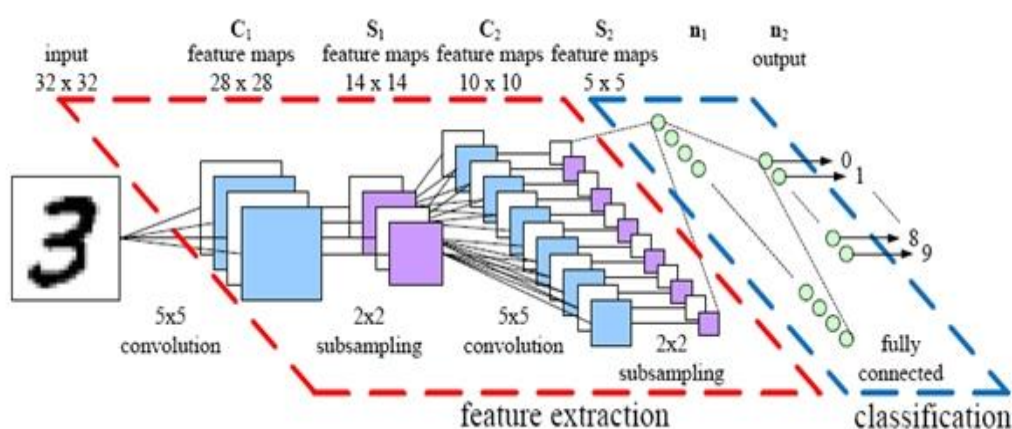


การนำนิรอลเน็ตเวิร์คมาใช้เป็นตัวจำแนก จะใช้ข้อมูลนำเข้าที่เป็นลำดับแถว 1 มิติ เช่น การทำนายราคายางพารา อาจใช้คุณลักษณะสำคัญที่เป็นของราคายางพาราย้อนหลัง 15 วัน ที่มีส่วนซึ้นนำเข้าเท่ากับ 15 ตัว

การใช้นิรอลเน็ตเวิร์คมาจำแนกข้อมูล 2 มิติ เช่น ภาพ อาจใช้คุณลักษณะเป็นจุดภาพ หรือใช้ตัวสกัดคุณลักษณะ (Feature Extractor) สกัดจากข้อมูลภาพที่มี 2 มิติ ให้เป็นคุณลักษณะของสิ่งที่อยู่ในภาพ โดยตัวสกัดคุณลักษณะที่เกี่ยวกับภาพมีมากมาย เช่น Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF), Haar-like Feature เป็นต้น โดยตัวสกัดคุณลักษณะเหล่านี้ถูกออกแบบให้ใช้ในงานที่ต่างกัน แต่ก็สามารถใช้งานได้กับงานอื่นๆ เช่นเดียวกัน ซึ่งผลลัพธ์อาจทำได้ดีไม่พอเท่ากับงานที่ถูกออกแบบมาให้ใช้เพื่อการจำแนก (Classify) ในขั้นต่อไป ดังนั้น คอนโวลูชันนอลนิรอลเน็ตเวิร์คจึงมีแนวคิดว่าจะสร้างตัวสกัดคุณลักษณะ จากชุดข้อมูลสอน (Training data) ที่มี เพื่อที่ไม่ต้องออกแบบตัวสกัดคุณลักษณะที่เหมาะสมกับข้อมูลทุกครั้ง

คอนโวลูชันนอลนิรอลเน็ตเวิร์คเป็นวิธีการทางการเรียนรู้ของเครื่อง (Machine Learning) โดยพัฒนามาจากนิรอลเน็ตเวิร์ค

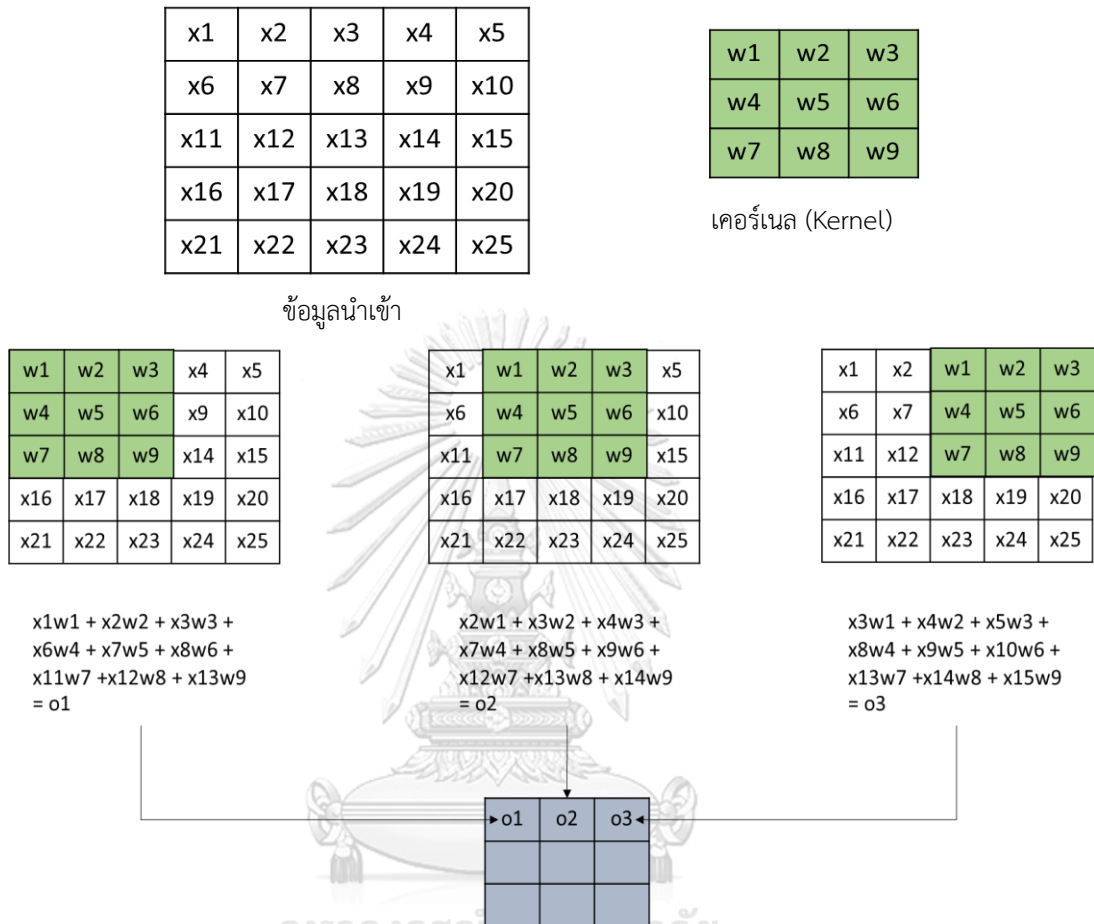
คอนโวลูชันนอลนิรอลเน็ตเวิร์คมีโครงสร้างย่อยคือ ชั้นการสกัดคุณลักษณะ (Trainable Feature Extractor Layer) และ ชั้นจำแนก (Classification Layer) ดังรูปที่ 2.2



รูปที่ 2.2 โครงสร้างของคอนโวลูชันนอลนิรอลเน็ตเวิร์ค [5]

ในขั้นของการสกัดคุณลักษณะประกอบไปด้วย ชั้นคอนโวลูชัน (Convolution) และชั้นสุ่มตัวอย่าง (Subsampling) โดยเคอร์เนลที่ใช้ในชั้นคอนโวลูชัน จะถูกสุ่มขึ้นมาในครั้งแรกเพื่อใช้ในการ

คำนวณชั้นคอนโวลูชัน วิธีการคอนโวลูชันจะนำข้อมูลนำเข้ามาทำการคอนโวลูท (Convolute) กับ เคอร์เนล แสดงดังรูปที่ 2.3



รูปที่ 2.3 วิธีการคอนโวลูชัน

หลังจากการทำคอนโวลูชัน ผลลัพธ์ที่ได้จะได้แผนที่ของคุณลักษณะ (Feature maps) ออกมาตามจำนวนเคอร์เนลที่ใช้ ซึ่งมีจำนวนมากขึ้นเพื่ออาจได้ข้อมูลเพิ่มเติม หลังจากนั้นนำเข้าสู่กระบวนการสุ่มตัวอย่าง (Pooling)

ในขั้นของสุ่มตัวอย่างนั้น ทำการลดจำนวนและความซ้ำซ้อนของข้อมูลจากชั้นของคอนโวลูชัน และในขั้นของการสุ่มตัวอย่าง วิธีที่เป็นที่นิยมคือการหาค่าสูงสุดหรือการหาค่าเฉลี่ย โดยขั้นของการสุ่มตัวอย่าง จะมีการเลื่อนวินโดว์ (Window) ที่ต่างจากชั้นคอนโวลูชัน คือไม่เลื่อนทับซ้อนกัน ต่อจากนั้นนำเข้าสู่ชั้นการจำแนกต่อไป โดยวิธีการของชั้นสุ่มตัวอย่างนั้นมีวิธีการทำดังรูปที่ 2.4

x1	x2	x3	x4
x5	x6	x7	x8
x9	x10	x11	x12
x13	x14	x15	x16

ผลลัพธ์จากชั้นคอนโวลูชัน



เคอร์เนล - ค่าเฉลี่ยการสุ่ม  
ตัวอย่าง

		x3	x4
		x7	x8
x9	x10	x11	x12
x13	x14	x15	x16

ค่าเฉลี่ยของ(x1,x2,x5,x6) = o1

x1	x2		
x5	x6		
x9	x10	x11	x12
x13	x14	x15	x16

ค่าเฉลี่ยของ(x3,x4,x7,x8) = o2

x1	x2	x3	x4
x5	x6	x7	x8
x9	x10		
x13	x14		

ค่าเฉลี่ย(x11,x12,x15,x16) = o4

o1	o2
o3	o4

รูปที่ 2.4 ตัวอย่างกระบวนการการสุ่มตัวอย่าง

หลังจากเข้าสู่ชั้นการสุ่มตัวอย่าง ผลลัพธ์ที่ได้ออกมา จะได้คุณลักษณะที่สามารถนำไปเข้าตัวจำแนกเพื่อแยกเป็นคลาสต่างๆ ซึ่งตัวจำแนกมีมากมายหลายแบบควรเลือกให้เหมาะสมกับข้อมูลที่น่าไปใช้ต่อไป ซึ่งผู้วิจัยได้เลือกใช้ตัวจำแนกแบบซอฟต์แวร์แมกซ์ (Softmax) โดยให้ผลลัพธ์เป็นค่าความน่าจะเป็นที่อยู่ระหว่าง 0-1 ดังสมการที่ 1

$$f(x) = \frac{e^x}{\sum e^x} \tag{1}$$

โดยที่  $x$  คือผลลัพธ์ของคุณลักษณะ และ  $f(x)$  คือ ค่าความน่าจะเป็นของแต่ละคลาส

หากผลลัพธ์บิดเบือนจากข้อมูลค่าความจริง (Ground truth) น้้าหนักของเคอร์เนล ในส่วนของตัวจำแนก และในชั้นของตัวสกัดคุณลักษณะ จะถูกปรับให้มีความเหมาะสมกับชุดข้อมูลสอน โดยวิธีการแพร่กระจายย้อนกลับ (Back propagation)

โดยในวิทยานิพนธ์ฉบับนี้นำวิธีการคอนโวลูชันนอลนิรอลเน็ตเวิร์คไปใช้ในขั้นตอนการหาบริเวณแพ่งด้านบนของรถประจำทาง ขั้นตอนการหาบริเวณที่มีข้อความ และขั้นตอนการรู้จำข้อความ

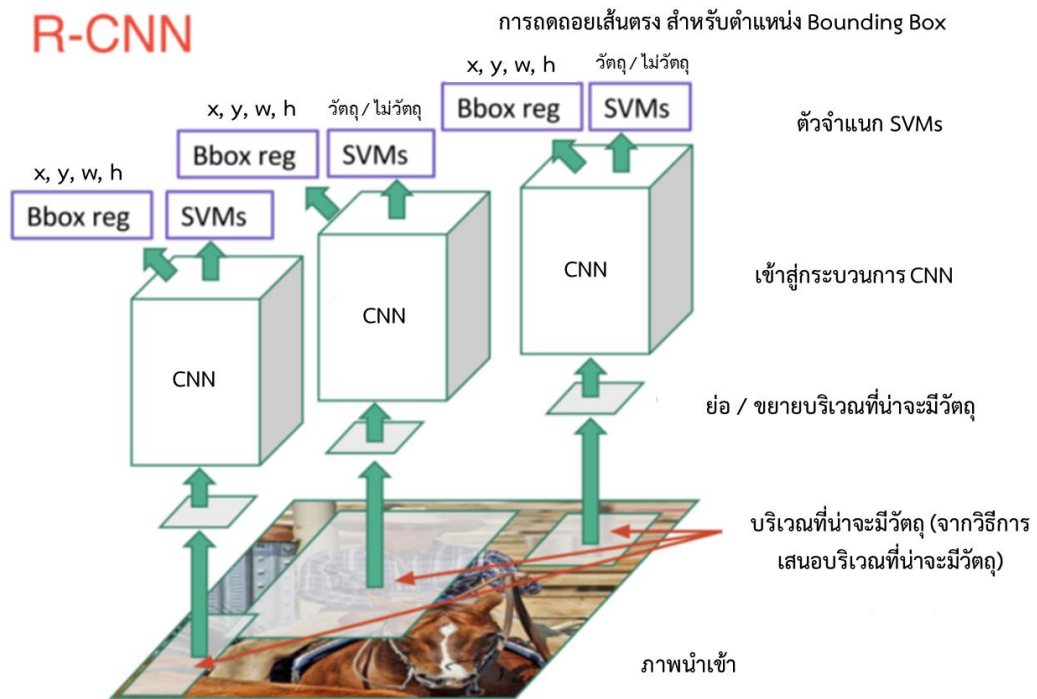
### 2.1.2 การตรวจจับวัตถุโดยวิธีการเสนอบริเวณที่น่าจะเป็นวัตถุ (Object Detection Proposal based Detector) [6]

ในปัจจุบันการตรวจจับวัตถุ มีวิธีการที่เสนอบริเวณที่น่าจะมีวัตถุ (Object Proposal Algorithm) ทำให้ไม่ต้องทำการเลื่อนวินโดว์ (Sliding Windows) หาบริเวณที่มีวัตถุทั้งภาพ เช่น ซีเล็กทีฟเสิช (Selective Search) [7] วิธีการเสนอบริเวณที่น่าจะมีวัตถุเหล่านี้ถูกสร้างโดยใช้ความรู้จากการมองเห็นของมนุษย์ เช่น สี เส้น คุณลักษณะต่างๆ เพื่อเสนอว่าบริเวณใดมีวัตถุ ทำให้สามารถลดบริเวณที่ต้องหาวัตถุได้มาก ซึ่งจากนั้นนำไปเข้ากระบวนการจำแนกต่อไป โดยวิธีการเสนอบริเวณที่น่าจะมีวัตถุถูกพัฒนาจนมีวิธีการอาร์ซีเอ็นเอ็น (Regions with Convolutional Neural Networks: R-CNN) และวิธีการ ฟาส อาร์ซีเอ็นเอ็น (Fast R-CNN) ซึ่งเป็นวิธีที่พัฒนาก่อนหน้า ของวิธีฟาสเตอร์อาร์ซีเอ็นเอ็น (Faster R-CNN)

#### 2.1.2.1 อาร์ซีเอ็นเอ็น (Regions with Convolutional Neural Networks: R-CNN) [8]

วิธีการอาร์ซีเอ็นเอ็น จะเริ่มจากใช้วิธีการเสนอบริเวณที่น่าจะมีวัตถุหาบริเวณที่น่าจะมีวัตถุ (เช่น ซีเล็กทีฟเสิช) จากนั้นทำการตัด ย่อ ขยาย โดยไม่สนใจอัตราส่วนภาพ (Aspect Ratio) โดยย่อขนาดให้เท่ากับความต้องการของตัวจำแนก จากนั้นนำเข้าสู่กระบวนการคอนโวลูชันนอลนิรอลเน็ตเวิร์ค ซึ่งเป็นตัวสกัดคุณลักษณะ เพื่อสร้างคุณลักษณะออกมา สุดท้ายนำไปเข้าตัวจำแนกเอสวิเอ็ม (SVM) เพื่อทำการจำแนกว่าบริเวณนั้นมีวัตถุหรือไม่ โดยวิธีการอาร์ซีเอ็นเอ็น แสดงดังรูปที่ 2.5

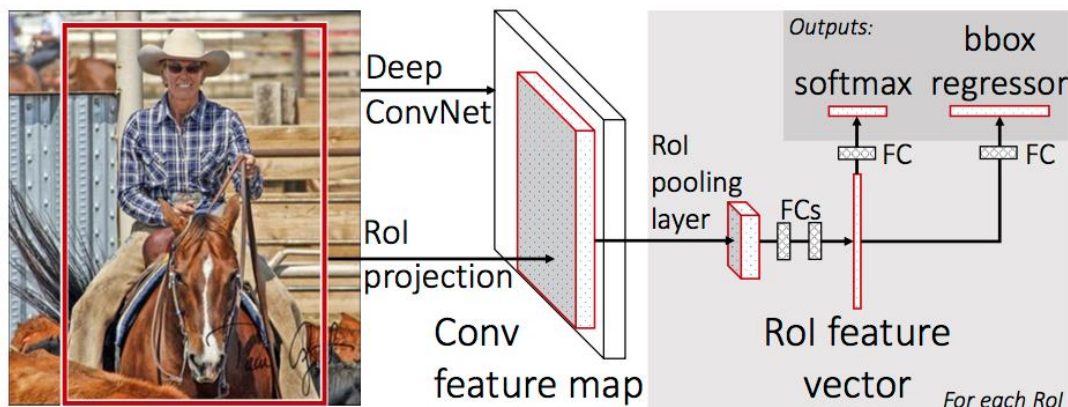
ข้อเสียของกระบวนการอาร์ซีเอ็นเอ็น คือ บริเวณที่น่าจะมีวัตถุ 1 บริเวณ จะต้องทำการเข้ากระบวนการคอนโวลูชันนอลนิรอลเน็ตเวิร์ค และจำแนกด้วยเอสวิเอ็ม ทีละ 1 บริเวณ และเนื่องจากบริเวณที่น่าจะมีวัตถุนั้นมีจำนวนมาก ทำให้ใช้เวลาและทรัพยากรในการประมวลผลมาก จึงได้มีการพัฒนาวิธีการฟาส อาร์ซีเอ็นเอ็น (Fast R-CNN) เข้ามาแก้ไข



รูปที่ 2.5 กระบวนการอาร์ซีเอ็นเอ็น [9]

### 2.1.2.2 ฟาส อาร์ซีเอ็นเอ็น (Fast R-CNN) [10]

ฟาส อาร์ซีเอ็นเอ็น เป็นวิธีที่พัฒนามาจาก อาร์ซีเอ็นเอ็น โดยเริ่มต้นจากนำภาพนำเข้าไปหาบริเวณที่น่าจะมีวัตถุจากวิธีการเสนอบริเวณที่น่าจะมีวัตถุ (เช่น ซีเล็คทีฟเลิซ) จากนั้นนำภาพนำเข้าทั้งภาพเข้าสู่กระบวนการคอนโวลูชันนอลนิวรอลเน็ตเวิร์คที่ถูกสอนจากตัวถ่ายโอนการเรียนรู้ จากนั้นจะใช้แผนภาพคุณลักษณะในขั้นสุดท้ายของการทำคอนโวลูชัน มาเปรียบเทียบกับตำแหน่งที่น่าจะมีวัตถุซึ่งได้จากวิธีการเสนอบริเวณที่มีวัตถุในตอนแรก โดยหาว่าตำแหน่งที่น่าจะมีวัตถุอยู่บริเวณใดในแผนภาพคุณลักษณะขั้นสุดท้ายของการทำคอนโวลูชัน หลังจากนั้นเข้าสู่กระบวนการ อาร์โอไอพูลลิง (ROI Pooling) เพื่อให้ได้ค่าคุณลักษณะที่ขนาดคงที่ (Fixed Size Feature Vector) จากนั้นเข้าสู่กระบวนการจำแนก ที่เป็นกระบวนการเชื่อมโยงเต็มรูปแบบ (Fully Connected Layer) และค่าซอฟต์แวร์แมกซ์ (Softmax) โดยสุดท้ายได้ผลลัพธ์ที่เป็นค่าของหมวดหมู่ของวัตถุ และจากตำแหน่งที่น่าจะมีวัตถุนำมาเข้ากระบวนการเชื่อมโยงเต็มรูปแบบอีกครั้ง เพื่อแก้ไขพิกัดของวัตถุที่จะตอบออกมาโดยวิธีการของฟาสอาร์ซีเอ็นเอ็น แสดงดังรูปที่ 2.6



รูปที่ 2.6 กระบวนการฟาส อาร์ซีเอ็นเอ็น [10]

โดยแต่ละตำแหน่งจะมีค่าพิกัดที่น่าจะมีวัตถุ ค่าพิกัดจะถูกนำมาคำนวณหาตำแหน่งบนแผนภาพคุณลักษณะขั้นสุดท้ายของการทำคอนโวลูชัน และนำเข้าสู่กระบวนการ อาร์โอไอพุอลิ่ง ทำให้วิธีฟาส อาร์ซีเอ็นเอ็น นำภาพนำเข้าสู่กระบวนการคอนโวลูชันนอลนิวรอลเน็ตเวิร์คเพียงครั้งเดียว ซึ่งสามารถลดระยะเวลาและการประมวลผลได้มาก เมื่อเปรียบเทียบกับวิธีการอาร์ซีเอ็นเอ็น ที่นำตำแหน่งที่น่าจะมีวัตถุ เข้าสู่กระบวนการคอนโวลูชันนอลนิวรอลเน็ตเวิร์คทุกบริเวณ

อย่างไรก็ตามวิธีการฟาส อาร์ซีเอ็นเอ็น ยังมีข้อเสียในเรื่องของการใช้วิธีการเสนอบริเวณที่น่าจะมีวัตถุที่ใช้เวลานานและเป็นส่วนที่ไม่เกี่ยวข้องกับวิธีการคอนโวลูชันนอลนิวรอลเน็ตเวิร์คใดๆ วิธีการฟาสเตอร์ อาร์ซีเอ็นเอ็น (Faster R-CNN) จึงออกแบบมาเพื่อพัฒนาในส่วนของวิธีการเสนอบริเวณที่สามารถทำการสอนด้วยวิธีการคอนโวลูชันนอลนิวรอลเน็ตเวิร์ค

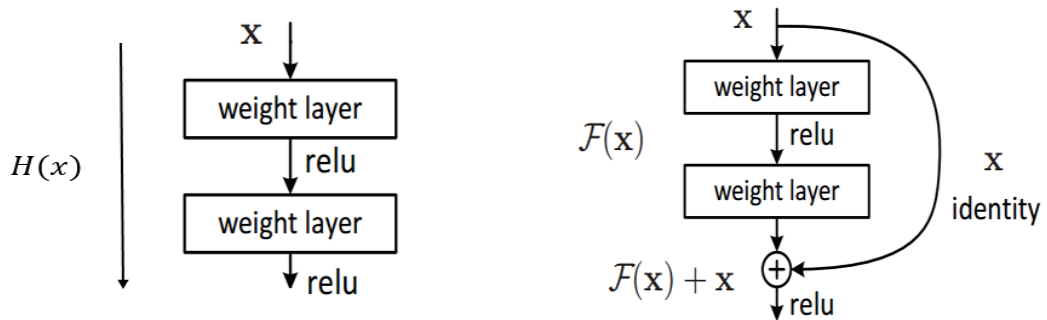
จุฬาลงกรณ์มหาวิทยาลัย

### 2.1.2.3 ฟาสเตอร์ อาร์ซีเอ็นเอ็น (Faster R-CNN) [11]

ฟาสเตอร์อาร์ซีเอ็นเอ็น (Faster RCNN) เป็นวิธีการเรียนรู้ของเครื่องบนพื้นฐานของ คอนโวลูชันนอลนิวรอลเน็ตเวิร์ค ที่นิยมใช้ในการตรวจจับวัตถุ (Object Detection) เนื่องจากผลลัพธ์มีความแม่นยำที่สูง และพัฒนามาจากฟาส อาร์ซีเอ็นเอ็น โดยโครงสร้างของฟาสเตอร์อาร์ซีเอ็นเอ็น แสดงดังรูปที่ 2.7



เรสสิดู เน็ตเวิร์ค (Residual Networks) หรือ เรสเน็ต (ResNets) [15] เกิดจากโครงสร้างที่ถูกเรียกว่า เรสสิดูบล็อก (Residual Block) นำมาต่อกันหลายๆ ชั้น จนเกิดเป็นโครงสร้างดังรูปที่ 2.8 โดย เรสเน็ต สามารถนำมาใช้ในการทำการสอนการเรียนรู้เชิงลึก และสามารถมีได้มากกว่า 100 ชั้น



**รูปที่ 2.9** เปรียบเทียบโครงสร้างคอนโวลูชันนอลนิวโรลเน็ตเวิร์ค ซ้าย: โครงสร้างคอนโวลูชันนอลนิวโรลเน็ตเวิร์คทั่วไป ขวา: โครงสร้างคอนโวลูชันนอลนิวโรลเน็ตเวิร์คแบบเรสสิดูบล็อก (Residual Block) [16]

จากรูปที่ 2.9 การทำงานของโครงสร้างคอนโวลูชันนอลนิวโรลเน็ตเวิร์คทั่วไป ทำตามลำดับดังรูปทางซ้าย เริ่มจากข้อมูล  $x$  และส่งข้อมูล  $x$  ไปยังส่วนของคอนโวลูชันนอลนิวโรลเน็ตเวิร์ค ( $H(x)$ ) จนกระทั่งได้ข้อมูลส่งออก

สำหรับรูปทางด้านขวาเป็นโครงสร้างที่มีการดัดแปลงมาจากโครงสร้างย่อยทั่วไป โดยมีการเพิ่มการคัดลอกข้อมูล  $x$  ในตอนแรก และนำไปรวมกับข้อมูลที่ได้จากคอนโวลูชันนอลนิวโรลเน็ตเวิร์ค ( $F(x)$ ) และทำการเข้าฟังก์ชันเรกติไฟต์เชิงเส้น (ReLU) ในตอนสุดท้าย โดยเรียกโครงสร้างทางขวาว่าเรสสิดูบล็อก

ฟังก์ชันเรกติไฟต์เชิงเส้น (Rectified Linear Unit : ReLU) เป็นฟังก์ชันกระตุ้นที่เป็นที่นิยมใช้ในหลายปีที่ผ่านมา ฟังก์ชันเรกติไฟต์เชิงเส้นให้ผลลัพธ์ออกมาเป็นจำนวนบวก หรือจำนวนที่มากกว่าเท่ากับศูนย์เสมอ โดยสามารถคำนวณได้จากสมการ 2

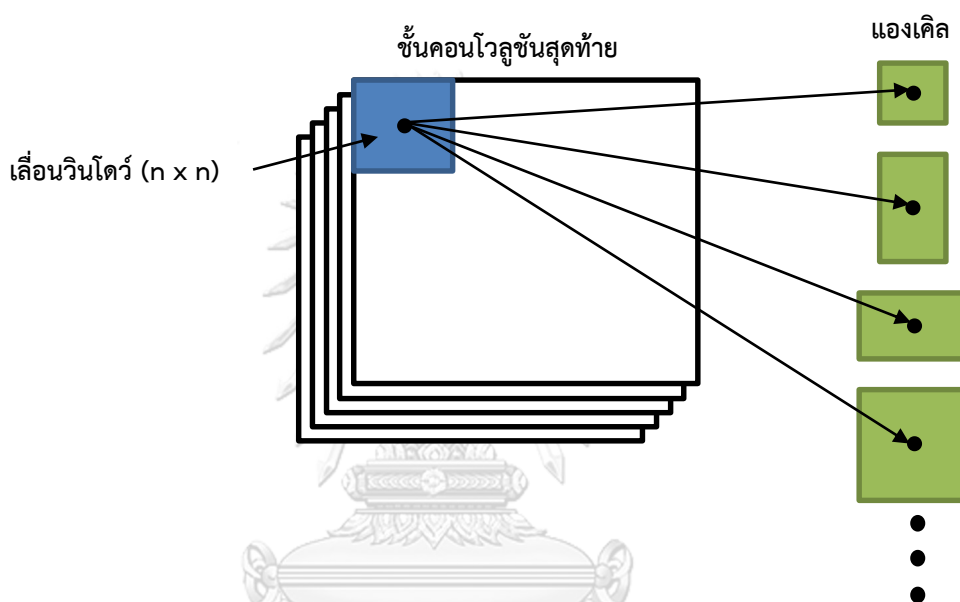
$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (2)$$

เมื่อเข้าสู่กระบวนการฟาสเตอร์อาร์ซีเอ็นเอ็น เริ่มจากภาพนำเข้ามายังตัวสกัดคุณลักษณะที่ใช้การถ่ายโอนการเรียนรู้จนได้ผลลัพธ์ของชั้นคอนโวลูชันสุดท้าย



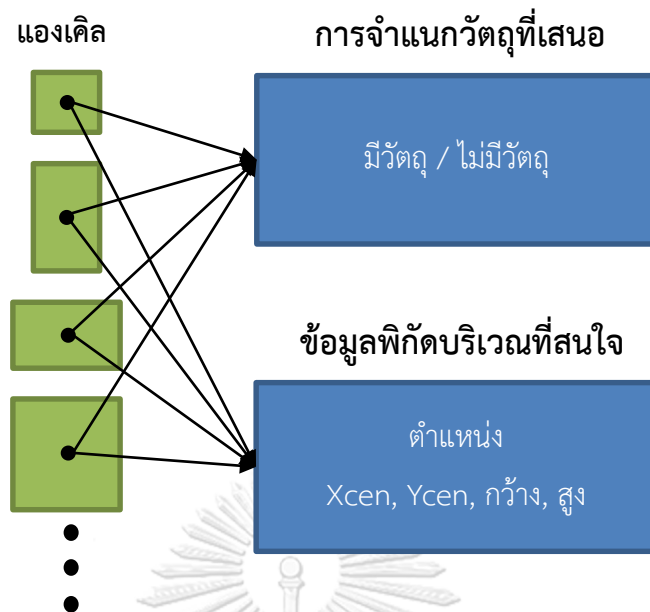
ตัวสกัดคุณลักษณะ สกัดได้ผลลัพธ์สุดท้ายจากชั้นคอนโวลูชันออกมาแล้ว นำส่งเข้าสู่กระบวนการเครือข่ายบริเวณที่เสนอ (Region Proposal Network : RPN) และกระบวนการอาร์ไอโอพูลลิง (ROI Pooling)

หลังจากเข้าสู่กระบวนการเครือข่ายบริเวณที่เสนอ จะมีการเลื่อนวินโดว์ที่มีขนาด  $n \times n$  บนชั้นคอนโวลูชันสุดท้าย แต่ละวินโดว์ที่ทำการเลื่อนไปบนชั้นคอนโวลูชันสุดท้าย จะมีเซตของกรอบสี่เหลี่ยมที่ล้อมรอบบริเวณที่น่าจะมีวัตถุอยู่ กรอบสี่เหลี่ยมเหล่านี้ เรียกว่า แอ็งเคิล (Anchors) โดยแสดงดังรูปที่ 2.10

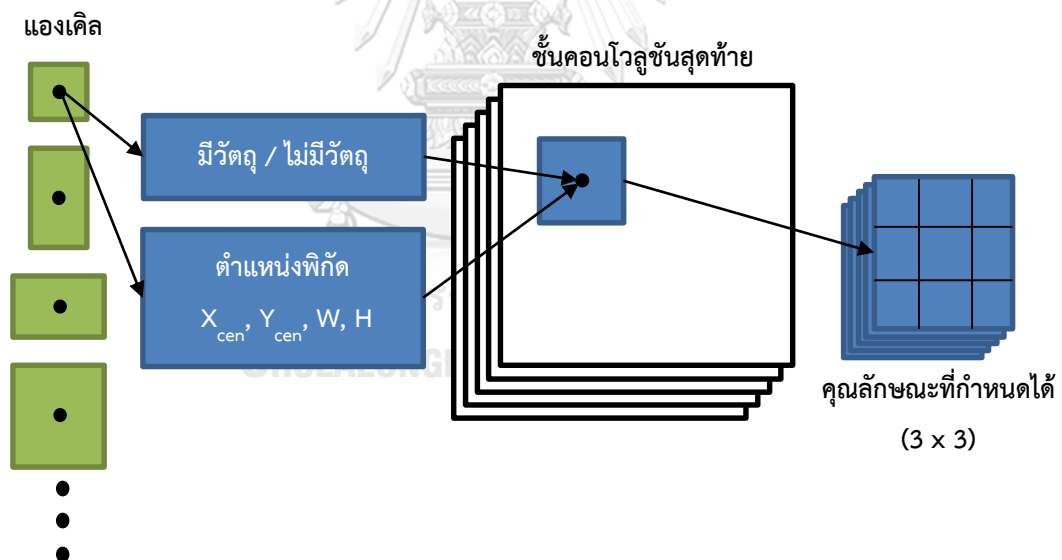


รูปที่ 2.10 กระบวนการเครือข่ายบริเวณที่เสนอ

แอ็งเคิลในวินโดว์ที่ถูกเลื่อนบนชั้นคอนโวลูชันสุดท้าย มีหลากหลายขนาด และหลากหลายอัตราส่วน โดยแต่ละแอ็งเคิลจะมีข้อมูลอยู่ 2 จำพวก จำพวกแรกคือค่าความเป็นวัตถุ ดูจากในแต่ละแอ็งเคิลว่ามีวัตถุหรือไม่ และยังใช้ค่าความเป็นวัตถุนี้ในการคัดกรองแอ็งเคิลให้เหลือเฉพาะแอ็งเคิลที่น่าจะเป็นวัตถุที่สุด (Object Proposal Classifier) จำพวกที่สองคือ ข้อมูลพิกัดจุดศูนย์กลาง ความสูง และความกว้างที่ล้อมรอบบริเวณที่น่าจะเป็นวัตถุของแต่ละแอ็งเคิล โดยข้อมูลจะสอดคล้องกับพิกัดของภาพนำเข้า (Region Proposal Regressor) โดยแสดงดังรูปที่ 2.11



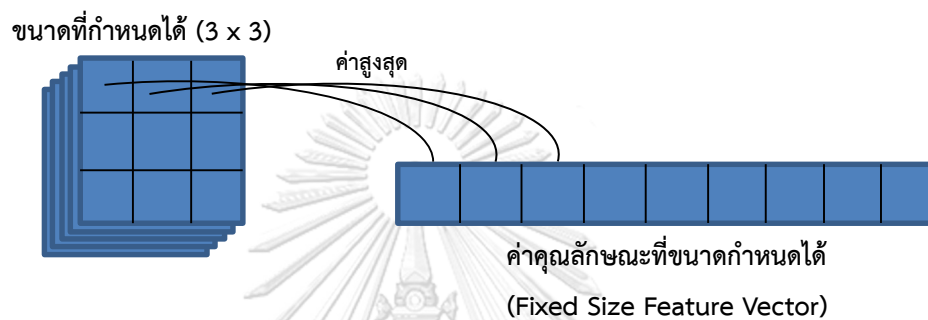
รูปที่ 2.11 กระบวนการเครือข่ายบริเวณที่เสนอ โดยข้อมูลของแต่ละแองเคิล มีข้อมูลการจำแนกวัตถุที่เสนอ และข้อมูลพิกัดบริเวณที่สนใจ



รูปที่ 2.12 กระบวนการอาร์โอไอ พูลลิง (ขั้นต้น)

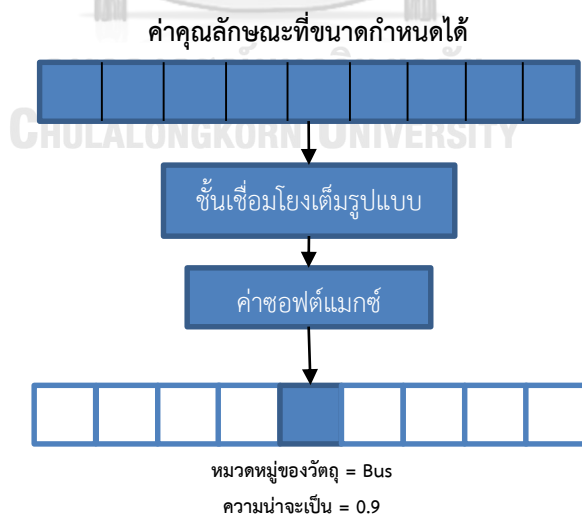
กระบวนการอาร์โอไอพูลลิง นำผลที่ได้จากกระบวนการเครือข่ายบริเวณที่เสนอว่า เป็นวัตถุหรือไม่ และอยู่ในตำแหน่งใด กับชั้นคอนโวลูชันสุดท้ายของตัวสกัดคุณลักษณะที่ถูกสอนมาใช้ข้อมูลร่วมกัน แต่เนื่องจากชั้นคอนโวลูชันสุดท้ายมีขนาดเล็กกว่าขนาดของภาพนำเข้า จึงไม่สามารถนำผลพิกัดที่ได้จากกระบวนการเครือข่ายบริเวณที่เสนอมาใช้ได้ทันที ต้องมีการปรับให้ข้อมูลตำแหน่งพิกัดให้สัมพันธ์กับตำแหน่งชั้นคอนโวลูชัน หลังจากนั้นเข้าสู่กระบวนการอาร์โอไอ พูลลิง โดยหาตำแหน่งที่

สัมพันธ์กันบนชั้นคอนโวลูชันสุดท้ายกับตำแหน่งพิกัดที่ได้จากกระบวนการเครือข่ายบริเวณที่เสนอ เมื่อได้บริเวณตามพิกัดบนชั้นคอนโวลูชันสุดท้าย จากนั้นทำการแบ่งข้อมูลของชั้นคอนโวลูชันสุดท้าย บริเวณที่ได้จากตำแหน่งพิกัดให้เป็นค่าขนาดคงที่ (Fixed Size) ยกตัวอย่างเช่น  $3 \times 3$  ดังรูปที่ 2.12 หลังจากนั้นทำการหาค่าที่มากที่สุดในแต่ละช่อง โดยสุดท้ายข้อมูลถูกส่งเก็บไปที่ ค่าคุณลักษณะที่ขนาดคงที่ (Fixed Size Feature Vector) ดังรูปที่ 2.13 เพื่อนำเข้าสู่กระบวนการจำแนก (Classification)



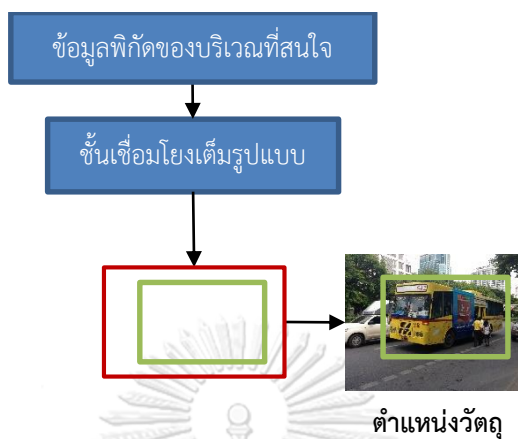
รูปที่ 2.13 กระบวนการอาร์ไอไอ พูลลิง (ชั้นหลัง)

กระบวนการจำแนก เป็นกระบวนการเชื่อมโยงเต็มรูปแบบ (Fully Connected Layer) และค่าซอฟต์แวร์แมกซ์ (Softmax) โดยสุดท้าย ได้ผลลัพธ์ที่เป็นค่าของหมวดหมู่ของวัตถุ ดังรูปที่ 2.14



รูปที่ 2.14 กระบวนการจำแนก (ผลลัพธ์ที่เป็นค่าของหมวดหมู่ของวัตถุ)

ตำแหน่งพิกัดของวัตถุนั้น นำมาเข้าขั้นตอนการเชื่อมโยงเต็มรูปแบบอีกครั้ง เพื่อแก้ไขพิกัดให้ใกล้เคียงกับบริเวณที่มีวัตถุมากที่สุด ดังรูปที่ 2.15



รูปที่ 2.15 กระบวนการจำแนก (แก้ไขพิกัดให้ใกล้เคียงกับบริเวณที่มีวัตถุ)

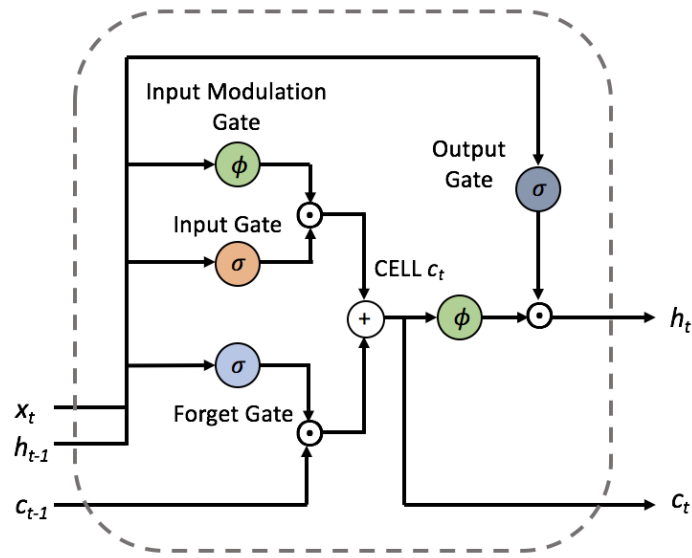
หลังจากคำนวณว่ามีวัตถุและตำแหน่งพิกัดในแต่ละแองเกิลจนครบ จะได้คำตอบที่แสดงให้เห็นถึงบริเวณที่มีวัตถุพร้อมตำแหน่งพิกัด โดยมีจำนวนมากตามจำนวนของแองเกิล จึงต้องใช้วิธีการขจัดค่าที่ไม่มากที่สุด (Non-Maximum Suppression : NMS) เพื่อผสานตำแหน่งพิกัดของกรอบของบริเวณที่เป็นวัตถุมากที่สุดให้เหลือเพียงหนึ่งเดียว

ในวิทยานิพนธ์ฉบับนี้ นำวิธีการคอนโวลูชันอลเน็ตเวิร์ค แบบฟาสเตอร์อาร์ซีเอ็นเอ็นไปใช้ในขั้นตอนการตรวจจับบริเวณแผงด้านบนของรถประจำทาง และขั้นตอนการหาบริเวณที่มีข้อความ

### 2.1.3 หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory: LSTM) [17]

หน่วยความจำระยะสั้นระยะยาว จะใช้งานกับข้อมูลที่มีลักษณะเป็นลำดับ (Sequence Data) ตัวอย่างเช่นอ่านหนังสือ ซึ่งเป็นข้อมูลที่มีลักษณะเป็นลำดับต่อกัน นำสิ่งที่อ่านไปแล้ว (Previous Hidden State) บวกกับสิ่งใหม่ที่ถูกอ่าน (Input Data) ทำให้เข้าใจความหมายในส่วนตรงที่อ่านได้

หน่วยความจำระยะสั้นระยะยาว ถูกสร้างขึ้นมาเพื่อแก้ปัญหาข้อมูลที่เป็นลักษณะลำดับที่ยาวมากๆ โดยทำหน้าที่เป็นตัวจดจำข้อมูลก่อนหน้าที่ถูกอ่านมาแล้ว มาวิเคราะห์กับข้อมูลที่ถูกอ่านปัจจุบัน โดยมีรายละเอียดดังรูปที่ 2.16



รูปที่ 2.16 การทำงานของหน่วยความจำระยะสั้นระยะยาว [18]

การทำงานของหน่วยความจำระยะสั้นระยะยาว สามารถบอกได้ว่า เมื่อไรที่ควรเขียน เมื่อไรที่ควรลืมหือเมื่อไรที่อนุญาตให้อ่านได้ โดย CELL เป็นตัวเก็บจำข้อมูลในหน่วยความจำระยะสั้นระยะยาว, Gate เป็นตัวที่ควบคุมการไหลของข้อมูล,  $x_t$  คือ ข้อมูลนำเข้า ณ เวลา  $t$ ,  $h_t$  คือ ข้อมูลส่งออก,  $h_{t-1}$  คือ ข้อมูลส่งออกก่อนหน้า,  $c_t$  คือ ข้อมูลที่ถูกเก็บจำ และ  $c_{t-1}$  คือ ข้อมูลที่ถูกเก็บจำก่อนหน้า

ประตูลืมหือ (Forget Gate :  $f_t$ ) คือการลบข้อมูลเดิมที่จำอยู่ออกไปเพื่อทำการเตรียมรับข้อมูลใหม่ ซึ่งหากประตูลืมหือมีค่าเป็น 0 จะลบข้อมูล แต่หากเป็น 1 ข้อมูลยังคงเดิม โดยดูข้อมูลนำเข้า ( $x_t$ ) ประกอบกับข้อมูลนำออกก่อนหน้านี ( $h_{t-1}$ ) และใช้ฟังก์ชันซิกมอยด์ (Sigmoid Function) เป็นตัวตัดสินใจ

ประตูนำเข้า (Input Gate :  $i_t$ ) หากมีข้อมูลใหม่เข้า ควรอัปเดตข้อมูลที่ตัวเก็บจำข้อมูลหรือไม่ ซึ่งหากประตูนำเข้ามีค่าเป็น 0 หมายถึงไม่ควรอัปเดตข้อมูล แต่หากเป็น 1 หมายถึงควรอัปเดตข้อมูล โดยดูข้อมูลนำเข้า ( $x_t$ ) ประกอบกับข้อมูลนำออกก่อนหน้านี ( $h_{t-1}$ ) และใช้ฟังก์ชันซิกมอยด์เป็นตัวตัดสินใจ

ประตูปรับค่านำเข้า (Input Modulation Gate :  $g_t$ ) คือ ค่าที่ไปอัปเดต หากค่าที่ประตูนำเข้าเป็น 1 โดยดูข้อมูลนำเข้า ( $x_t$ ) ประกอบกับข้อมูลนำออกก่อนหน้านี ( $h_{t-1}$ ) และใช้ฟังก์ชันแทนเฮซ ( $\tanh$ )

การอัปเดตข้อมูลที่ตัวเก็บจำ ( $c_t$ ) คือ หลังจากได้ข้อมูลจากประตูลืมหือ ประตูนำเข้า และประตูปรับค่านำเข้า จะสามารถอัปเดตข้อมูลที่ตัวเก็บจำได้ ดังสมการที่ 3

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (3)$$

จากสมการที่ 3 หากประตูลืมมีค่าเป็น 0 จะไม่นำข้อมูลที่ถูเก็บจำเดิม ( $C_{t-1}$ ) มาพิจารณา แต่หากประตูลืมมีค่าเป็น 1 จะนำข้อมูลที่ถูเก็บจำเดิม ( $C_{t-1}$ ) มาพิจารณา และหากประตูนำเข้ามีค่าเป็น 1 จะนำค่าจากประตูปรับค่านำเข้า ( $g_t$ ) มาอัปเดต และได้ค่า  $C_t$  ค่าใหม่

ประตูนำออก (Output Gate :  $O_t$ ) คือเป็นตัวตัดสินใจว่าจะอนุญาตให้ ข้อมูลนำออก ( $h_t$ ) สามารถออกเพื่อนำไปพิจารณาในลำดับถัดไปได้หรือไม่ โดยดูข้อมูลนำเข้า ( $X_t$ ) ประกอบกับข้อมูลนำออกก่อนหน้านี้ ( $h_{t-1}$ ) และใช้ฟังก์ชันซิกมอยด์เป็นตัวตัดสินใจ

ข้อมูลนำออก ( $h_t$ ) จะนำข้อมูลที่ถูเก็บจำ ( $C_t$ ) เข้าสู่ฟังก์ชันแทนเอช แล้วนำไปพิจารณากับ ประตูนำออก ว่าสามารถให้ข้อมูลนำออก ออกไปได้หรือไม่ หากประตูนำออก มีค่าเป็น 0 ข้อมูลนำออกจะไม่สามารถออกไปได้ และหากค่าเป็น 1 จะอนุญาตให้ข้อมูลนำออกออกไปได้ ดังสมการที่ 4

$$h_t = o_t \odot \tanh(c_t) \quad (4)$$

ในวิทยานิพนธ์ฉบับนี้นำเอาเรื่องของหน่วยความจำระยะสั้นระยะยาว มาใช้ในขั้นตอนการรู้จำข้อความ

#### 2.1.4 สูตรฮาเวอร์ไซน์ (Haversine Formula) [19]

สูตรฮาเวอร์ไซน์ คือสูตรที่ใช้คำนวณระยะทางจริงบนผิวโลก โดยใช้ค่าละติจูด และลองจิจูดของ 2 ตำแหน่งที่ต่างกัน โดยสูตรฮาเวอร์ไซน์ ค้นพบโดยศาสตราจารย์เจมส์ อินมาน ในปี ค.ศ.1835 และเป็นที่นิยมใช้ในระบบสารสนเทศทางภูมิศาสตร์ (Geographic Information System : GIS)

ระยะทางระหว่าง ละติจูด และลองจิจูด 2 ตำแหน่งที่ต่างกัน สามารถคำนวณได้จากสูตรฮาเวอร์ไซน์ ดังต่อไปนี้

$$a = \sin^2\left(\frac{(\varphi_2 - \varphi_1) \times \frac{\pi}{180}}{2}\right) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2\left(\frac{(\lambda_2 - \lambda_1) \times \frac{\pi}{180}}{2}\right) \quad (5)$$

$$c = 2 \cdot \operatorname{atan2}(\sqrt{a}, \sqrt{1-a}) \quad (6)$$

$$d = R \cdot c \quad (7)$$

โดยที่  $\varphi_1$  และ  $\varphi_2$  คือ ละติจูดตำแหน่งที่ 1 และที่ 2,  $\lambda_1$  และ  $\lambda_2$  คือ ลองจิจูดของตำแหน่งที่ 1 และที่ 2,  $R$  คือ รัศมีของโลก = 6,371 กิโลเมตร และ  $d$  คือ ระยะที่คำนวณได้ โดยการคำนวณระยะระหว่างละติจูด และลองจิจูด 2 ตำแหน่งจะเริ่มจากสมการที่ 5 6 และ 7 ตามลำดับ

การคำนวณเริ่มจากสมการที่ 5 โดยนำพิกัดละติจูดและลองจิจูดทั้ง 2 ตำแหน่งมาลบกัน และแปลงเป็นหน่วยเรเดียนโดยการคูณด้วย  $\frac{\pi}{180}$  จากนั้นแทนค่าตัวแปรที่เหลือ และคำนวณออกมาจนได้ค่า  $a$

เมื่อได้ค่า  $a$  เรียบร้อยแล้ว จะนำมาแทนค่าในสมการที่ 6 โดยมีการใช้ฟังก์ชันคืนค่าแทนเจนต์ตรีโกณมิติผกผัน ของจุดพิกัดแกน  $x$  และ  $y$  ที่ระบุ หรือ  $\text{atan2}(x, y)$  [20] โดยพิกัดที่  $x$  แทนด้วย  $\sqrt{a}$  และ  $y$  แทนด้วย  $\sqrt{(1-a)}$  สุดท้ายคำนวณตามสมการจนได้ค่า  $c$  และเข้าสู่สมการที่ 7 เพื่อคำนวณหาระยะทาง  $d$

ในวิทยานิพนธ์ฉบับนี้ นำสูตรของฮาวเวอร์ไซน์มาใช้ในขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ โดยใช้ในการคำนวณระยะทางของตำแหน่งละติจูด ลองจิจูด ของป้ายรถประจำทาง กับตำแหน่งที่ถ่ายภาพ

### 2.1.5 เทนเซอร์โฟล (TensorFlow) [21]

การพัฒนาทางด้านการเรียนรู้ของเครื่อง อาจมีความยากในการเขียนโปรแกรมเพื่อพัฒนา เนื่องจากมีกระบวนการที่ซับซ้อน จึงทำให้เกิดการพัฒนาไลบรารีที่สามารถช่วยให้พัฒนาทางด้านการเรียนรู้ของเครื่อง ทางกุเกิดจึงได้พัฒนาไลบรารีที่ช่วยในการพัฒนาทางด้านการเรียนรู้ขึ้นมา เรียกว่า เทนเซอร์โฟล (TensorFlow)

เทนเซอร์โฟล เป็นไลบรารีโอเพนซอร์ส และใช้ภาษาไพธอน จาวา ซีพลัสพลัส ในการเขียน โดยเทนเซอร์โฟล ถูกนำมาใช้ในกระบวนการการสร้างตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง

### 2.1.6 ระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด (Minimum Editing Distance) [22]

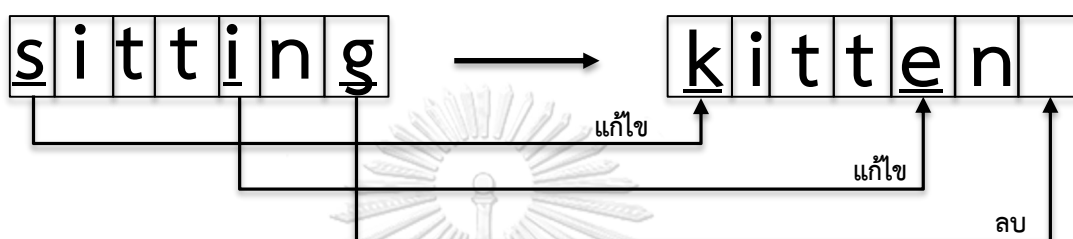
เป็นการเปรียบเทียบคำตอบกับข้อมูลชุดคำตอบ ว่าคำตอบนั้นใกล้เคียงกับชุดข้อมูลคำตอบตัวใด โดยใช้ฟังก์ชันของ `difflib.get_close_matches` [23] ซึ่งใช้หลักการของ ระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด (Minimum Editing Distance)

ระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด จะเป็นการเปรียบเทียบสตริง (String) จำนวน 2 ชุด ว่าต้องทำการเพิ่ม (Insertion) ลบ (Deletion) หรือแทนที่ (Substitution)

ตัวอักษรตำแหน่งใด จึงจะทำให้มีตัวอักษรที่เหมือนกัน โดยการทำการเพิ่ม ลบ หรือแทนที่ จะมีค่าน้ำหนักเท่ากับ 1

s i t t i n g → k i t t e n

(ก)



(ข)

รูปที่ 2.17 ตัวอย่างการแก้ไขด้วยจำนวนน้อยที่สุด (ก) ตัวอย่างสตริง 2 ชุดที่ต้องการแก้ไข โดยจะแก้ไขคำจากทางซ้าย ไปเป็นคำทางด้านขวา (ข) กระบวนการทำการแก้ไขจากตัวอย่างสตริงในรูป (ก)

รูปที่ 2.17 แสดงตัวอย่างการแก้ไขสตริง 2 ชุด จากคำว่า sitting เป็นคำว่า kitten โดยจะต้องทำการเพิ่ม ลบ หรือแทนที่ด้วยจำนวนกี่ครั้ง รูปที่ 2.17 (ข) แสดงถึงกระบวนการที่ต้องทำการแก้ไขจำนวน 2 ครั้ง และลบ 1 ครั้ง รวมแล้วจะมีค่าน้ำหนักในการแก้ไขที่น้อยที่สุดเท่ากับ 3 ในการแก้ไข sitting เป็นคำว่า kitten

การคำนวณการแก้ไขด้วยจำนวนน้อยที่สุดนั้น สามารถคำนวณได้จากโปรแกรมดังต่อไปนี้

```
for (int i=0; i<=string1.length(); i++) { //เริ่มจากตัวอักษรที่ 0 ใน string1 แล้ววนลูป
    for (int j=0; j<=string2.length(); j++) { //เริ่มจากตัวอักษรที่ 0 ใน string2 แล้ววนลูป
        //หาก String1 ไม่มีค่าใดๆ จะทำการ Insertion ทุกตัวอักษรของ String2
        if (i==0)
            valueTable[i][j] = j
        //หาก String2 ไม่มีค่าใดๆ จะทำการ Insertion ทุกตัวอักษรของ String1
        else if (j==0)
            valueTable[i][j] = i
```



```

//หากตัวอักษรในตำแหน่งที่ i ของ String1 และตำแหน่ง j ของ String2 ตรงกัน ค่า
valueTable จะนำมาจาก valueTable[i-1][j-1]

else if (string1 [i] == string2 [j])
    valueTable[i][j] = valueTable[i-1][j-1]
//หากตัวอักษรแตกต่างกัน จะหาค่า valueTable ที่ตำแหน่ง [i][j-1], [i-1][j-1], [i-1][j] ที่
มีค่าน้อยที่สุด และนำมาบวก 1 เพื่อเป็นค่า valueTable ที่ตำแหน่ง [i][j]

else
    valueTable[i][j] = min(valueTable[i][j-1], valueTable[i-1][j-1],
        valueTable[i-1][j]) + 1
    }
}

```

รูปที่ 2.18 แสดงถึงโปรแกรมการคำนวณการแก้ไขด้วยจำนวนน้อยที่สุด

จากรูปที่ 2.18 จะเป็นขั้นตอนการคำนวณการแก้ไขด้วยจำนวนที่น้อยที่สุด โดยที่  $i$  คือตัวอักษรของ String1,  $j$  คือตัวอักษรของ String2 และ valueTable คือค่านำหนักการเปลี่ยนแปลงที่ได้ ผลลัพธ์ที่ได้จะแสดงดังตารางที่ 2.1 เมื่อคำนวณจนเสร็จสิ้น ผลลัพธ์ในตำแหน่งสุดท้ายที่คำนวณได้จะเป็นผลลัพธ์ที่บอกถึงค่านำหนักที่ต้องใช้ในการแก้ไขจากคำว่า sitting เป็น kitten ซึ่งมีค่าเท่ากับ 3

ตารางที่ 2.1 แสดงตัวอย่างการคำนวณวิธีการแก้ไขที่น้อยที่สุด

		String 2 (j)							
		0	1	2	3	4	5	6	
			k	i	t	t	e	n	
String 1 (i)	0		0	1	2	3	4	5	6
	1	s	1	1	2	3	4	5	6
	2	i	2	2	<u>1</u>	2	3	4	5
	3	t	3	3	2	<u>1</u>	2	3	4
	4	t	4	4	3	2	<u>1</u>	2	3
	5	i	5	5	4	3	2	2	3
	6	n	6	6	5	4	3	3	<u>2</u>
	7	g	7	7	6	5	4	4	<b>3</b>

```
>>> get_close_matches('appel', ['ape', 'apple', 'peach', 'puppy'])
['apple', 'ape']
```

รูปที่ 2.19 กระบวนการจับคู่กันของลำดับ [23]

โดยฟังก์ชัน *difflib.get\_close\_matches* ดูจากค่าที่มีการแก้ไขที่น้อยที่สุด แล้วจะนำคำตอบที่มีการแก้ไขที่น้อยที่สุด มาตอบเป็นคำตอบลำดับแรก

จากรูปที่ 2.19 คำที่ต้องการให้จับคู่คือคำว่า ‘appel’ และชุดข้อมูลของคำตอบที่เป็นไปได้ทั้งหมดคือ ‘ape’, ‘apple’, ‘peach’, ‘puppy’ โดยคำตอบที่ได้จะตอบคำตอบที่ใกล้เคียงที่สุดกับคำว่า ‘appel’ นั่นคือ ‘apple’ ซึ่งมีคะแนนความน่าจะเป็นที่ใกล้เคียงมากที่สุด โดยมีคำตอบที่คะแนนความน่าจะเป็นที่ใกล้เคียงที่สุดถัดมาเป็นคำว่า ‘ape’

โดยระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด ถูกนำมาใช้ในส่วนของการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ในการเปรียบเทียบเลขสายรถประจำทางที่ได้กับข้อมูลจีพีเอสป้ายรถประจำทาง

### 2.1.7 การวัดประสิทธิภาพ (Performance Evaluation)

การวัดประสิทธิภาพของระบบอ่านเลขสายรถประจำทาง ประกอบไปด้วยการวัดประสิทธิภาพ 3 ชนิด คือ การวัดค่าไอโอยู การวัดประสิทธิภาพโดยใช้คอนฟิวชันเมตริกซ์ และ การวัดประสิทธิภาพจำแนกโดยคลาส

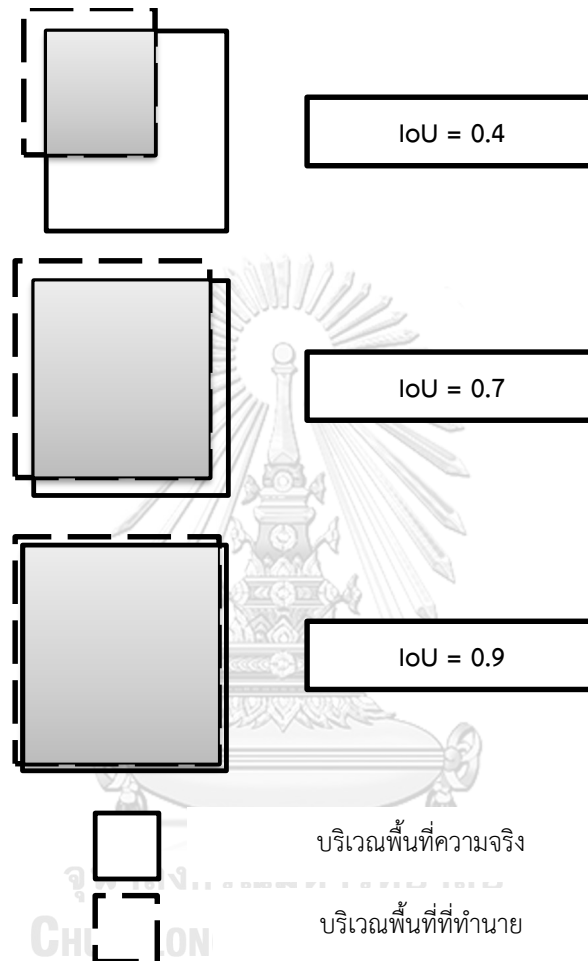
#### 2.1.7.1 ค่าไอโอยู (Intersection over Union: IoU) [24]

ค่าไอโอยู เป็นค่าที่นิยมใช้ในการวัดค่าความแม่นยำ (Accuracy) ของการตรวจจับวัตถุ ซึ่งในการแข่งขันปาสคาล วิโอซี (PASCAL VOC) [25] ซึ่งเป็นการแข่งขันการตรวจจับวัตถุที่มีชื่อเสียง ได้ใช้ค่าไอโอยูในการประเมินคะแนนการแข่งขัน

การวัดค่าไอโอยู สามารถวัดได้จากอัตราส่วนพื้นที่ที่ทับซ้อน (Area of Overlap) ระหว่างบริเวณที่ทำนาย (Prediction Area) กับ บริเวณค่าความจริง (Ground Truth Area)หารด้วย พื้นที่รวมทั้งหมด (Area of Union) ระหว่างบริเวณที่ทำนาย และบริเวณค่าความจริง โดยแสดงดังสมการที่ 8

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (8)$$

ในงานวิจัยนี้ใช้ค่าไอโอยู ในการประเมินความถูกต้องของการตรวจจับบริเวณแพงด้านบนของรถประจำทาง หากค่าไอโอยูมีค่าเกินกว่า 0.7 ถือว่าตรวจจับเจอบริเวณแพงด้านบนของรถประจำทาง โดยตัวอย่างของการวัดผลค่าไอโอยู สามารถแสดงได้ดังรูปที่ 2.18



รูปที่ 2.20 แสดงตัวอย่างของการวัดผลค่าไอโอยู

#### 2.1.7.2 คอนฟิวชันเมทริกซ์ (Confusion Matrix)

คอนฟิวชันเมทริกซ์ คือการประเมินผลลัพธ์จากการทำนาย หรือผลลัพธ์ที่ได้จากโปรแกรม โดยเปรียบเทียบกับผลลัพธ์จริง

ตารางที่ 2.2 แสดงคอนฟิวชันเมทริกซ์การจำแนกบริเวณแผงด้านบนของรถประจำทาง

		คลาสจริง	
		แผงด้านบนของรถประจำทาง	ไม่ใช่ แผงด้านบนของรถประจำทาง
คลาสที่ทำนาย	แผงด้านบนของรถประจำทาง	TP	FP
	ไม่ใช่ แผงด้านบนของรถประจำทาง	FN	TN

คอนฟิวชันเมทริกซ์ของการจำแนกข้อมูลประเภท 2 คลาส เช่น ในระบบของวิทยานิพนธ์ฉบับนี้ซึ่งประกอบด้วย 2 คลาส คือ บริเวณส่วนภาพแผงด้านบนของรถประจำทาง กับ ไม่ใช่บริเวณแผงด้านบนของรถประจำทาง ค่าในตารางกำหนดให้มีค่าดังนี้

1) True Positive (TP) คือ ส่วนภาพที่ระบบทำนายว่าเป็นบริเวณแผงด้านบนของรถประจำทาง และส่วนภาพนั้นเป็นบริเวณแผงด้านบนของรถประจำทางจริง (มีผลลัพธ์ของค่าไอโออยู่มากกว่าเท่ากับ 0.7)

2) False Positive (FP) คือ ส่วนภาพที่ระบบทำนายว่าเป็นบริเวณแผงด้านบนของรถประจำทาง แต่ส่วนภาพนั้น ไม่ใช่บริเวณแผงด้านบนของรถประจำทาง (มีผลลัพธ์ของค่าไอโออยู่น้อยกว่า 0.7)

3) True Negative (TN) คือ ส่วนภาพที่ระบบทำนายว่า ไม่ใช่บริเวณแผงด้านบนของรถประจำทาง และส่วนภาพนั้น ไม่ใช่บริเวณแผงด้านบนของรถประจำทาง

4) False Negative (FN) คือ ส่วนภาพที่ระบบทำนายว่า ไม่ใช่บริเวณแผงด้านบนของรถประจำทาง แต่ส่วนภาพนั้นเป็นบริเวณแผงด้านบนของรถประจำทาง (มีผลลัพธ์ของค่าไอโออยู่น้อยกว่า 0.7)

### 2.1.7.3 ตัววัดประสิทธิภาพการจำแนกโดยคลาส [26]

การวัดประสิทธิภาพการจำแนก สามารถคำนวณหาค่าความเที่ยง (Precision) ค่ารีคอล (Recall) และค่าการวัดเอฟ (F-Measure) ได้จากสมการดังต่อไปนี้

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

ค่าความเที่ยง (Precision) คือค่าที่บอกว่าจากผลการทำนายว่าเป็นบริเวณแผงหน้ารถประจำทางนั้น ถูกต้องกับค่าจริงร้อยละเท่าไร สามารถคำนวณได้จากอัตราส่วนของข้อมูลที่ทำนายบริเวณแผงป้ายด้านบนของรถประจำทางได้ถูกต้อง (TP) กับข้อมูลที่ทำนายว่าเป็นบริเวณแผงป้ายด้านบนของรถประจำทางทั้งหมด (TP+FP)

ค่ารีคอล (Recall) คือค่าที่บอกว่าจากจำนวนบริเวณแผงหน้ารถประจำทางจริงทั้งหมดนั้น ระบบทำนายถูกต้องร้อยละเท่าไร สามารถคำนวณจาก อัตราส่วนของข้อมูลที่ทำนายบริเวณแผงป้ายด้านบนของรถประจำทางได้ถูกต้อง (TP) กับข้อมูลที่เป็นบริเวณแผงป้ายด้านบนของรถประจำทางทั้งหมด (TP+FN)

ค่าการวัดเอฟ (F-measure) คือ การวัดประสิทธิภาพโดยรวมของทั้งสองค่า ระหว่างค่าความเที่ยง และค่ารีคอล ซึ่งนำค่าทั้งสองมาคำนวณร่วมกัน โดยสามารถคำนวณได้จาก อัตราส่วนของ สองเท่าของค่าความเที่ยงคูณกับค่ารีคอล กับ ค่าความเที่ยงบวกกับค่ารีคอล

## 2.2 งานวิจัยที่เกี่ยวข้อง

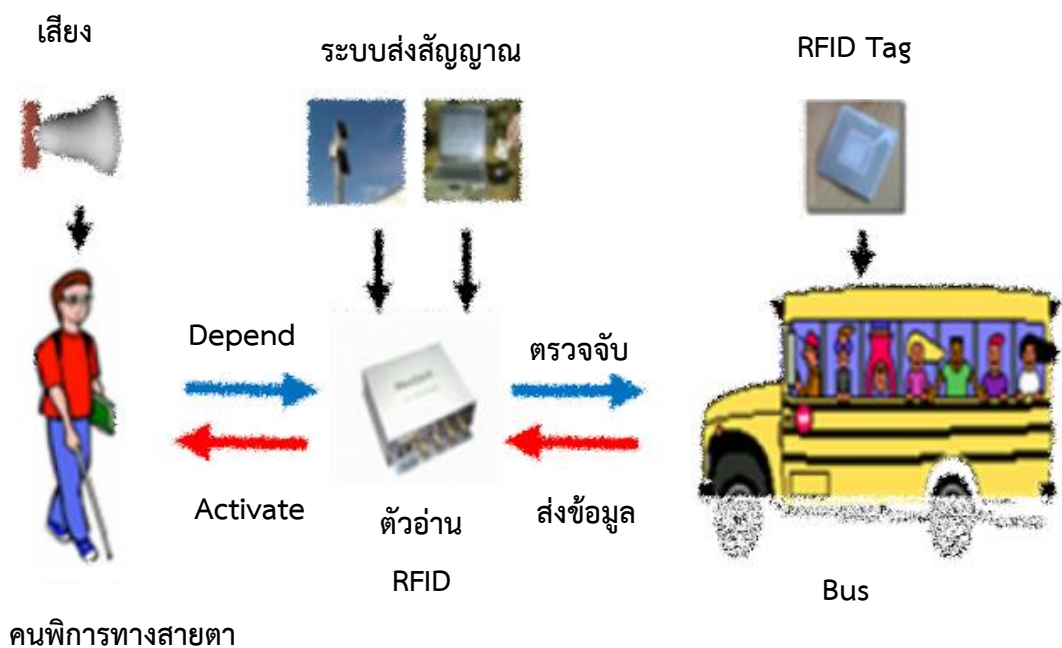
ผู้วิจัยได้ทำการสำรวจงานวิจัยที่เกี่ยวข้องกับการรู้จำเลขหมายรถประจำทาง โดยแบ่งออกเป็น 2 รูปแบบคือ งานวิจัยการตรวจหาเลขสายรถประจำทางโดยใช้เครื่องรับรู้ (Sensor) และงานวิจัยการตรวจหารถประจำทางและเลขสายรถประจำทางโดยใช้วิธีการประมวลผลภาพ

### 2.2.1 การตรวจหาเลขสายรถประจำทางโดยใช้เครื่องรับรู้

การตรวจหาเลขสายรถประจำทางโดยใช้เครื่องรับรู้อาร์เอฟไอดี (Radio Frequency Identification : RFID) เป็นเทคโนโลยีที่รับส่งข้อมูลสัญญาณผ่านคลื่นวิทยุ โดยมีหลักการคล้ายกับการอ่านบาร์โค้ด แต่เครื่องรับรู้อาร์เอฟไอดี สามารถรับส่งสัญญาณระยะไกลได้ตามความถี่ของตัวรับและตัวส่ง จึงมีการนำมาประยุกต์ใช้กับรถประจำทาง โดยมีตัวรับ ตัวส่งสัญญาณติดตั้งที่รถประจำทางและที่ป้ายรถประจำทาง เครื่องรับรู้อาร์เอฟไอดีทำงานร่วมกับระบบฐานข้อมูลเพื่อให้ทราบข้อมูลของรถประจำทางคันดังกล่าว

Noor และคณะ [27] ได้นำเสนอระบบตรวจจับรถประจำทางสำหรับผู้พิการทางสายตา โดยใช้ระบบเครื่องรับรู้อาร์เอฟไอดี ในการอ่านหรือรับรู้ว่ารถประจำทางสายใดกำลังเข้าป้าย จากนั้นทำการส่งเคราะห์เสียง เพื่อแจ้งข้อมูลของรถประจำทางให้ผู้โดยสารที่อยู่ที่ป้ายรถประจำทางทราบ

Kalbani และคณะ [28] ได้นำเสนอระบบตรวจจับรถประจำทางโดยใช้เครื่องรับรูดอาร์เอฟไอดี เพื่อช่วยเหลือผู้พิการทางสายตา โดยแบ่งออกเป็น 2 ระบบ คือระบบสถานี และระบบบนรถประจำทาง ทั้ง 2 ระบบนี้เชื่อมต่อภายในฐานข้อมูลเดียวกัน โดยระบบสถานีทำการอ่านเครื่องรับรูดอาร์เอฟไอดีของรถประจำทางที่เข้าสถานีเพื่อหารหัสข้อมูลของรถประจำทางในฐานข้อมูล และมีการส่งเป็นเสียงเพื่อบ่งบอกให้ผู้พิการทางสายตาทราบ สำหรับระบบที่ตัวรถประจำทาง เมื่อเข้าใกล้กับสถานี ตัวอ่านเครื่องรับรูดอาร์เอฟไอดีบนรถประจำทางจะอ่านเครื่องรับรูดอาร์เอฟไอดีของสถานี โดยมีข้อมูลจากฐานข้อมูลบอกถึงชื่อของสถานี มีเสียงบอกเตือนผู้โดยสารในรถถึงชื่อสถานีที่จะถึง และบอกจำนวนผู้พิการทางสายตาที่ถือบัตรโดยสารที่มีเครื่องรับรูดอาร์เอฟไอดี แก่ผู้ขับรถประจำทาง โดยบัตรโดยสารนั้นมีข้อมูลจุดหมายปลายทางที่ผู้พิการทางสายตานั้นจะเดินทางไป และยังได้จัดทำเว็บไซต์สำหรับเจ้าหน้าที่เพื่อเข้าถึงระบบ โดยระบบทั้งหมดมีการทำงานดังรูปที่ 2.21



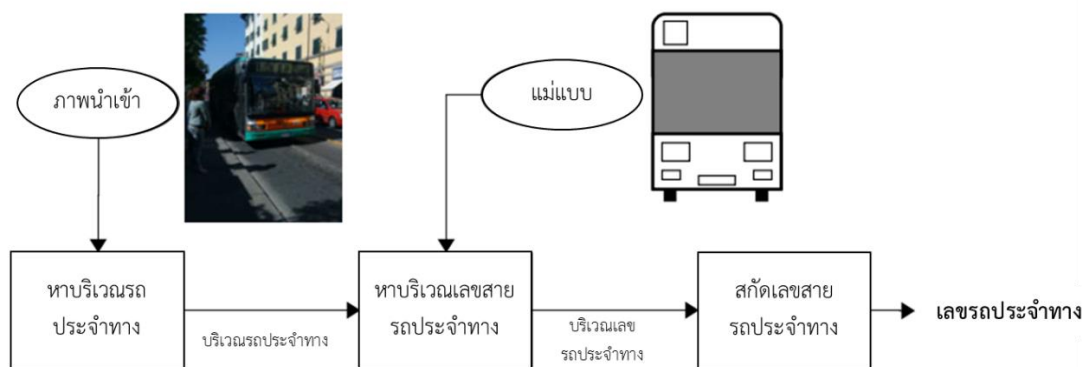
รูปที่ 2.21 การทำงานของระบบตรวจจับรถประจำทางโดยใช้เครื่องรับรูดอาร์เอฟไอดี [28]

งานวิจัยการตรวจหาเลขสายรถประจำทางโดยใช้เครื่องรับรูดตามที่สำรวจข้างต้น พบว่างานวิจัยที่ใช้ตัวรับรูดเป็นตัวดำเนินการ สามารถระบุข้อมูลของรถประจำทางได้ดีในระยะใกล้ที่เครื่องรับรูดอาร์เอฟไอดีสามารถทำงานได้เท่านั้น อีกทั้งมีค่าใช้จ่ายที่สูง เพราะต้องติดตั้งเครื่องรับรูดในรถประจำทาง และบริเวณป้ายรถประจำทางทุกตำแหน่ง

## 2.2.2 การตรวจหาเลขสายรถประจำทางโดยใช้วิธีการประมวลผลภาพ

การสกัดจากภาพถ่ายหรือภาพวิดีโอเพื่อหารถประจำทาง หรือ หาเลขสายรถประจำทาง ได้มีงานวิจัยที่ได้นำเสนอวิธีการสกัดโดยใช้คุณลักษณะต่างๆ จากภาพหรือวิดีโอที่ค้นเพื่อนำไปวิเคราะห์บริเวณรถประจำทาง และเลขสายรถประจำทาง

Guida และคณะ [29] ได้นำเสนอการตรวจจับรถประจำทางและการอ่านเลขสายรถประจำทาง โดยได้นำเอาอัลกอริทึมการเรียนรู้ของเครื่อง ของ Viola and Jones [30] ที่ออกแบบไว้สำหรับตรวจจับใบหน้า ประยุกต์ใช้กับภาพที่มีรถประจำทาง 1,500 ภาพ และภาพที่ไม่มีรถประจำทาง 2,100 ภาพ เพื่อตรวจจับว่ามีรถประจำทางในภาพหรือไม่ หลังจากนั้นเข้าสู่ขั้นตอนของการหาเลขสายรถประจำทาง โดยการสร้างแม่แบบ (Template) หลังจากนั้นนำภาพที่ตรวจจับว่ามีรถประจำทางนั้น มาใช้หลักการการแก้ไขภาพแบบเพอร์สเปคทีฟ เพื่อให้ได้หน้ารถประจำทางที่หน้าตรง และส่งต่อไปเปรียบเทียบกับแม่แบบที่ได้กำหนดไว้แล้วว่า เลขของสายรถประจำทางนั้นอยู่ตำแหน่งใด และหาบริเวณส่วนบนของรถซึ่งเป็นส่วนของป้ายบอกบอกข้อมูลต่างๆ ของรถประจำทาง โดยใช้วิธีการหาเส้นขอบแบบแคนนี่ (Canny edge detection) และฮัฟทรานส์ฟอร์ม (Hough transform) หลังจากนั้นทำการสกัดตัวเลขสายรถประจำทางโดยใช้ แบบจำลองสี HSV เพื่อตัดแยกเลขรถประจำทางที่มีสีเฉพาะออกมา และใช้หลักการของค่าขีดแบ่งแบบปรับได้ (Adaptive Threshold) ในการแปลงให้เป็นภาพไบนารี สุดท้ายใช้โปรแกรมประยุกต์ การรู้จำอักขระด้วยแสง (OCR) เพื่ออ่านตัวเลขที่สกัดได้ออกมาในรูปแบบของเสียง โดยขั้นตอนการทำงานของระบบจะแสดงดังรูปที่ 2.22



รูปที่ 2.22 ขั้นตอนการทำงานของ การตรวจจับรถประจำทางและการอ่านเลขสายรถประจำทาง [29]

Pan และคณะ [31] ได้นำเสนอระบบที่สามารถตรวจจับข้อมูลของรถประจำทาง ซึ่งถูกถ่ายจากกล้องที่ติดตั้งบริเวณป้ายรถประจำทางดังรูปที่ 2.23 โดยระบบ จะตรวจจับรถประจำทางที่วิ่งเข้าสู่ป้าย และมีเสียงบอกเตือนผู้พิการทางสายตาที่สถานี โดยเป็นข้อมูลของรถประจำทางที่เข้ามาบริเวณ

ป้ายรถประจำทาง ระบบจะแบ่งเป็น 2 ระบบย่อยคือ ระบบตรวจจับรถประจำทาง และระบบตรวจจับเลขสายรถประจำทาง ระบบตรวจจับรถประจำทางนั้น ใช้ตัวสกัดคุณลักษณะ Histogram of the Oriented Gradient (HOG) สกัดหน้ารถประจำทาง และใช้ตัวจำแนก Support Vector Machine (SVM) เพื่อตรวจจับหน้ารถประจำทาง จากนั้นระบบตรวจจับเลขสายรถประจำทางจะทำการตรวจจับสายเลขรถประจำทางและข้อความต่างๆ จากหน้าของรถประจำทาง และเนื่องจากตัวเลขตัวอักษรที่แสดงผลบริเวณป้ายบอกหมายเลขสายรถประจำทาง นั้นมีลักษณะเป็นจุด ต่อกันจนเป็นอักษร ทำให้ตรวจจับยากกว่าตัวอักษรที่อยู่บนเอกสารต่างๆ จึงใช้หลักการ Adjacent character grouping และใช้วิธีการหาเส้นขอบแบบแคนนี่ เพื่อตรวจหาเค้าโครงของตัวอักษรและใช้ตัวจำแนก อดาบู้สต์ (Adaboost) ในการจำแนกเพื่อหาตัวอักษร หลังจากนั้น ใช้ซอฟต์แวร์ของการรู้จำอักขระด้วยแสง (OCR) เพื่อแปลงเป็นตัวอักษรข้อความ และสุดท้ายแปลงให้เป็นสัญญาณเสียงออกมาให้ผู้พิการทางสายตาทราบ

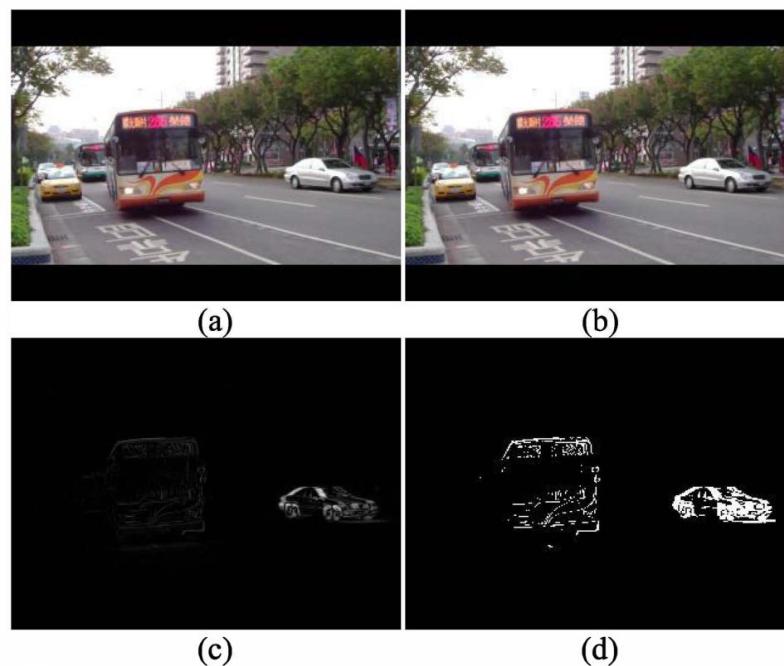


รูปที่ 2.23 รูปนำเข้าจากกล้องที่ติดตั้งบริเวณป้ายรถประจำทาง [31]

Tsai และคณะ [32] ได้นำเสนอระบบการตรวจจับตัวอักษรด้านหน้ารถประจำทางจากกล้องวิดีโอที่ติดตั้งบริเวณสถานี เพื่อช่วยเหลือผู้พิการทางสายตาให้สามารถรับรู้ถึงเลขสายรถประจำทางหรือข้อความที่อยู่บริเวณหน้ารถประจำทาง โดยระบบประกอบไปด้วย 3 ขั้นตอนคือ การตรวจจับรถประจำทางขณะเคลื่อนที่ การตรวจจับป้ายบอกหมายเลขสายรถประจำทาง และการตรวจจับบริเวณตัวอักษร สำหรับการตรวจจับรถประจำทางขณะเคลื่อนที่ ใช้หลักการ Modify Adaptive Frame Differencing (MAFD) ซึ่งเป็นวิธีการเปรียบเทียบความเปลี่ยนแปลงในความต่างของภาพ มาใช้ในการตรวจจับรถประจำทางขณะเคลื่อนที่ โดยเปรียบเทียบความแตกต่างระหว่างลำดับภาพก่อนหน้าและลำดับภาพปัจจุบัน ดังรูปที่ 2.24 หากมีความแตกต่างน้อยกว่าค่าขีดแบ่ง (Threshold) แสดงถึง



รถประจำทางไม่ได้เข้าใกล้กับกล้องตรวจจับบริเวณป้ายรถประจำทาง แต่หากมีความแตกต่างมากกว่าค่าขีดแบ่ง จะแสดงถึงรถประจำทางได้เข้าใกล้กับกล้องตรวจจับบริเวณป้ายรถประจำทาง เมื่อรถประจำทางเข้ามาใกล้ในระยะที่เหมาะสม รถประจำทางจะถูกตรวจจับและแสดงผลออกมาอยู่ในรูปของ Bounding Box หลังจากนั้นพิจารณาหาป้ายบอกหมายเลขสายรถประจำทาง โดยมีสมมติฐานว่าป้ายมักอยู่บริเวณด้านบนบนเสมอ และนำมาพิจารณาตรวจจับตัวอักษรโดยหาบริเวณที่มีตัวอักษรอยู่ ใช้หลักการ Background Distribution และ Border Thresholding เพื่อให้ได้บริเวณตัวอักษรจากป้ายที่แสดงข้อความหน้ารถประจำทาง และนำไปเข้าสู่กระบวนการอ่านข้อความต่อไป



รูปที่ 2.24 การตรวจจับรถประจำทาง (a) ภาพลำดับก่อนหน้า (b) ภาพลำดับปัจจุบัน (c) ภาพความต่างของ (a)(b), (d) ภาพไบนารีโดยจากค่าขีดแบ่ง [32]

งานวิจัยของ Dongjin Lee และคณะ [33] ได้นำเสนอระบบรู้จำตัวเลข สามารถนำไปช่วยเหลือผู้พิการทางการมองเห็นให้สามารถรู้ถึงเลขและข้อมูลของสายรถประจำทางจากภาพถ่ายทั่วไป โดยผลลัพธ์แสดงออกมาในรูปแบบของเสียง และหน้าจอบริเวณป้ายรถประจำทาง โดยหลักการของงานวิจัยนี้แบ่งเป็น 2 ขั้นตอน คือ การตรวจจับเลขสายรถประจำทาง และการรู้จำเลขสายรถประจำทาง โดยภาพนำเข้าจะเป็นภาพจากกล้องจำนวน 2 ตัวที่เชื่อมกันดังรูปที่ 2.25 ส่วนของการตรวจจับเลขสายรถประจำทางนั้นได้ทำการดัดแปลงวิธี Census Transform feature based AdaBoost ที่โดยปกติใช้สำหรับตรวจจับใบหน้ามนุษย์ มาทำการสอนและจำแนกตัวเลข ซึ่งตัวเลขที่นำมาสอนนั้นนำมาจากแผ่นป้ายทะเบียนรถ ลำดับถัดไปคือการรู้จำเลขสายรถประจำทาง ซึ่งแบ่งออกไปสามขั้นตอนย่อยคือ การเตรียมภาพ การสกัดคุณลักษณะ และการจำแนก โดยส่งไปยังขั้นตอน

สุดท้ายคือระบบการอ่านออกเสียง (Text to Speech) ในส่วนของการเตรียมภาพ ใช้หลักการ Otsu's thresholding และ Sauvola's thresholding มาช่วยกัน เนื่องจากสภาพแสงและเงาที่หลากหลายทำให้ตรวจจับยาก ส่วนการสกัดคุณลักษณะ และการจำแนก ใช้หลักการ 8-direction gradient feature ใช้หลักการนี้คำนวณแต่ละจุดภาพ จาก Sobel operator และนำมาจำแนก โดยใช้วิธี Multiclass Support vector machine with radial basis function โดยใช้ไลบรารีของ LIBSVM และสุดท้าย นำเอาระบบการอ่านออกเสียงจากกูเกิล มาใช้สังเคราะห์เสียง



รูปที่ 2.25 ภาพนำเข้าเป็นภาพจากกล้อง 2 ตัวที่เชื่อมกัน [33]

งานวิจัยของ Tsai และคณะ [34] ได้นำเสนอระบบตรวจจับเลขสายรถประจำทางจากภาพถ่ายวิดีโอที่ติดตั้งบริเวณป้ายรถประจำทาง เริ่มจากการเรียนรู้แยกสีโดยใช้ปริภูมิสีของ YCbCr ทำการเรียนรู้แบบมีผู้สอน โดยคำนวณจากค่าปริภูมิสี Y Cb และ Cr ที่เหมาะสมจากสีของเลขสายรถประจำทาง หลังจากนั้นทำการแปลงภาพให้อยู่ในรูปของภาพไบนารีโดยใช้ค่าขีดแบ่ง ดังรูปที่ 2.26 จากนั้นใช้วิธีขององค์ประกอบที่เชื่อมกัน (Connected Components) ในการหาตำแหน่งของเลขสายรถประจำทาง โดยแสดงผลในรูปของ Bounding Box แต่เนื่องจากมีจำนวน Bounding Box เป็นจำนวนมากที่สกัดมาจากวิธีองค์ประกอบที่เชื่อมกัน จึงได้เสนอวิธีการที่ทำการลบ Bounding Box ที่ไม่ใช่บริเวณที่เป็นเลขสายรถประจำทางออกไปโดยวิเคราะห์จากอัตราส่วนของ Bounding Box บริเวณที่มีเลขสายรถประจำทาง จนได้เฉพาะ Bounding Box ที่มีเลขสายรถประจำทาง โดยสามารถพัฒนาเพื่อรู้จำและแปลงเป็นเสียงเพื่อบ่งบอกได้ต่อไป



รูปที่ 2.26 ระบบตรวจจับเลขสายรถประจำทางจากภาพถ่ายวิดีโอ  
(a) ภาพลำดับนำเข้า (b) ภาพไบนารี [34]

Ma และคณะ [35] ได้นำเสนอระบบตรวจจับและรู้จำรถประจำทาง โดยใช้ภาพถ่ายจากกล้องรักษาความปลอดภัยที่ติดตั้งบริเวณถนน โดยระบบจะทำการสร้างโมเดล 3 มิติ จากภาพวิดีโอ เริ่มต้นที่การสร้างเส้นขอบรอบ (Contour) และทำการสร้างโมเดล 3 มิติจากเส้นขอบรอบ เพื่อหาส่วนที่เป็นรถประจำทาง หลังจากนั้นใช้วิธีการของ Line Segment Detector (LSD) เพื่อหาเส้นที่เหมาะสมกับเส้นของโมเดล 3 มิติ ดังรูปที่ 2.27 เพื่อทำการจับคู่ (Matching) โดยมีขั้นตอนคือ การทำมอร์โฟโลยี กับภาพที่ทำ LSD และนำไปทับซ้อนกับโมเดล 3 มิติ หากค่าการทับซ้อนระหว่าง LSD กับโมเดล 3 มิตินั้น มากกว่าค่าที่กำหนดไว้ ถือว่าเป็นรถประจำทาง โดยผลการทดลองสามารถตรวจจับหารรถประจำทางได้ดี

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.27 ระบบตรวจจับและรู้จำรถประจำทาง (a) ภาพลำดับนำเข้า (b) ภาพจากวิธีการ LSD [35]

งานวิจัยของ Cheng และคณะ [36] ได้นำเสนองานวิจัย การตรวจจับเลขสายรถประจำทาง จากภาพที่เลือกมาเฉพาะบริเวณที่เป็นเลขสายรถประจำทาง โดยวิเคราะห์เฉพาะส่วนที่เป็นสีแดงจาก ปริภูมิสี RGB หลังจากนั้นทำการเลือกให้ภาพพื้นหน้ามีสีแดง ซึ่งบริเวณสีแดงก็คือเลขสายรถประจำ ทาง หลังจากนั้นใช้วิธีการตัดแปลงจาก Otsu Threshold ในการขีดแบ่ง ผลลัพธ์ที่ได้เป็นภาพไบนารี ที่เห็นเลขสายรถประจำทาง ที่มีการตัดสิ่งรบกวน และลดการสะท้อนจากกระจกหน้าของรถประจำ ทาง ดังรูปที่ 2.28 ขั้นตอนต่อไป ใช้วิธีอิงค์ประกอบที่เชื่อมกัน (Connected Component) ในการ หาตำแหน่งของเลขสายรถประจำทาง โดยแสดงอยู่ในรูปของ Bounding Box ซึ่งเกิดสิ่งรบกวนที่ไม่ ต้องการเกิดขึ้นมากมาย จึงต้องทำการลบสิ่งรบกวนออกไป ให้เหลือเฉพาะบริเวณที่เป็นเลขสายรถ ประจำทางที่สนใจ เพื่อนำเลขสายรถประจำทางที่ค้นหามา ไปวิเคราะห์ต่อไป



รูปที่ 2.28 ระบบตรวจจับและรู้จำเลขสายรถประจำทาง (a) ภาพลำดับนำเข้า (b) เลือกภาพพื้นหน้า บริเวณรถประจำทางเป็นสีแดง (c) ภาพไบนารีหลังจากการทำ Otsu Threshold [36]

จากการสำรวจงานวิจัยดังที่กล่าวมา ผู้วิจัยประเมินว่าไม่สามารถนำมาใช้กับภาพถ่ายรถ ประจำทางในกรุงเทพมหานครได้ เนื่องจากในต่างประเทศ รูปแบบของรถประจำทางรวมทั้งป้ายบอก เลขสายรถประจำทางมีความเป็นมาตรฐาน และงานวิจัยที่กล่าวมาได้ใช้สมมุติฐานของมาตรฐานตัวรถ ประจำทาง และมาตรฐานป้ายรถประจำทางมาใช้ในการออกแบบขั้นตอนวิธีการหาเลขสายรถประจำ ทาง ส่วนรถประจำทางในพื้นที่ของกรุงเทพมหานคร และปริมณฑลนั้น มีหลากหลายรูปแบบมาก ทั้ง ลักษณะรถ สี ลักษณะป้าย ลักษณะตัวอักษร ลักษณะตัวเลข ฯลฯ รวมทั้งสภาพแวดล้อมที่ซับซ้อน และหลากหลาย

### 2.3 ประเภทของข้อมูลที่ใช้ในงานวิจัย

งานวิจัยนี้ใช้รูปถ่ายรถประจำทางของ องค์การขนส่งมวลชนกรุงเทพ (ขสมก.) และรถร่วม บริการ ที่ให้บริการในกรุงเทพมหานคร โดยกล่าวถึงข้อมูลบริเวณที่เป็นแผงด้านบนของรถประจำทาง

ประเภทของแผงด้านบนรถประจำทางที่แสดงข้อมูล ประเภทของข้อมูลที่แสดงบนแผงด้านบนรถประจำทาง และประเภทการแสดงผลข้อมูลบริเวณแผงด้านบนรถประจำทาง

### 2.3.1 บริเวณแผงด้านบนของรถประจำทาง

บริเวณแผงด้านบนของรถประจำทางคือ ส่วนบริเวณที่เป็นแผงป้ายด้านบนของด้านหน้ารถประจำทาง ที่แสดงถึงข้อมูลสายรถประจำทาง และอาจมีข้อมูลบอกว่าไปที่ใด ตัวอย่างบริเวณแผงด้านบนของรถประจำทาง แสดงดังรูปที่ 2.29



รูปที่ 2.29 บริเวณแผงด้านบนของรถประจำทาง ที่แสดงเฉพาะข้อมูลสายรถประจำทาง

### 2.3.2 ประเภทของแผงด้านบนรถประจำทางที่แสดงข้อมูล

บริเวณแผงด้านบนของรถประจำทาง แสดงถึงข้อมูลสายรถประจำทาง โดยรถประจำทางในกรุงเทพมหานครและปริมณฑล จะมีลักษณะแผงด้านบนรถประจำทางที่บอกข้อมูลของเลขสายรถประจำทางอยู่ 2 ประเภท คือ แผงด้านบนรถประจำทางที่ติดกับตัวรถ และแผงด้านบนรถประจำทางที่มีส่วนยื่นออกไปจากตัวรถ ดังรูปที่ 2.30 และรูปที่ 2.31



รูปที่ 2.30 แผงด้านบนรถประจำทางที่ติดกับตัวรถ



รูปที่ 2.31 แผงด้านบนรถประจำทางที่ยื่นออกมาจากตัวรถ

### 2.3.3 ประเภทของข้อมูลที่แสดงบนบริเวณแผงด้านบนของรถประจำทาง

ประเภทของข้อมูลที่แสดงบนบริเวณแผงด้านบนของรถประจำทางในกรุงเทพมหานครและปริมณฑล นั้น มีอยู่ 3 ประเภท คือ ประเภทที่มีเลขสายรถประจำทางอย่างเดียว ประเภทที่มีเลขสายรถประจำทางกับข้อมูลสถานที่ที่ผ่าน และ ประเภทที่มีเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ โดยแสดงดังรูปที่ 2.32 รูปที่ 2.33 และรูปที่ 2.34 ตามลำดับ



รูปที่ 2.32 รถประจำทางที่ข้อมูลบริเวณแผงด้านบนของรถประจำทางปรากฏเพียงตัวเลขสายรถประจำทาง



รูปที่ 2.33 รถประจำทางที่ข้อมูลบริเวณแผงด้านบนของรถประจำทางมีเลขสายรถประจำทางและข้อมูลสถานที่ที่ผ่าน



รูปที่ 2.34 รถประจำทางที่ข้อมูลบริเวณแผงด้านบนของรถประจำทางมีเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ

#### 2.3.4 ประเภทการแสดงผลข้อมูลบริเวณแผงด้านบนรถประจำทาง

การแสดงผลของข้อมูล ทั้งเลขสายรถประจำทาง หรือ ข้อมูลต่างๆ ของรถประจำทางที่บริการในกรุงเทพมหานครและปริมณฑล มีอยู่ 2 ประเภท คือ การแสดงผลแบบพิมพ์หรือทาสีถาวร และการแสดงผลแบบหลอดไฟแอลอีดี ดังรูปที่ 2.35 และรูปที่ 2.36 ตามลำดับ



รูปที่ 2.35 รถประจำทางที่แสดงข้อมูลบริเวณแผงด้านบนของรถประจำทางแบบ พิมพ์หรือทาสีถาวร



รูปที่ 2.36 รถประจำทางที่แสดงข้อมูลบริเวณแผงด้านบนของรถประจำทางแบบ หลอดไฟแอลอีดี

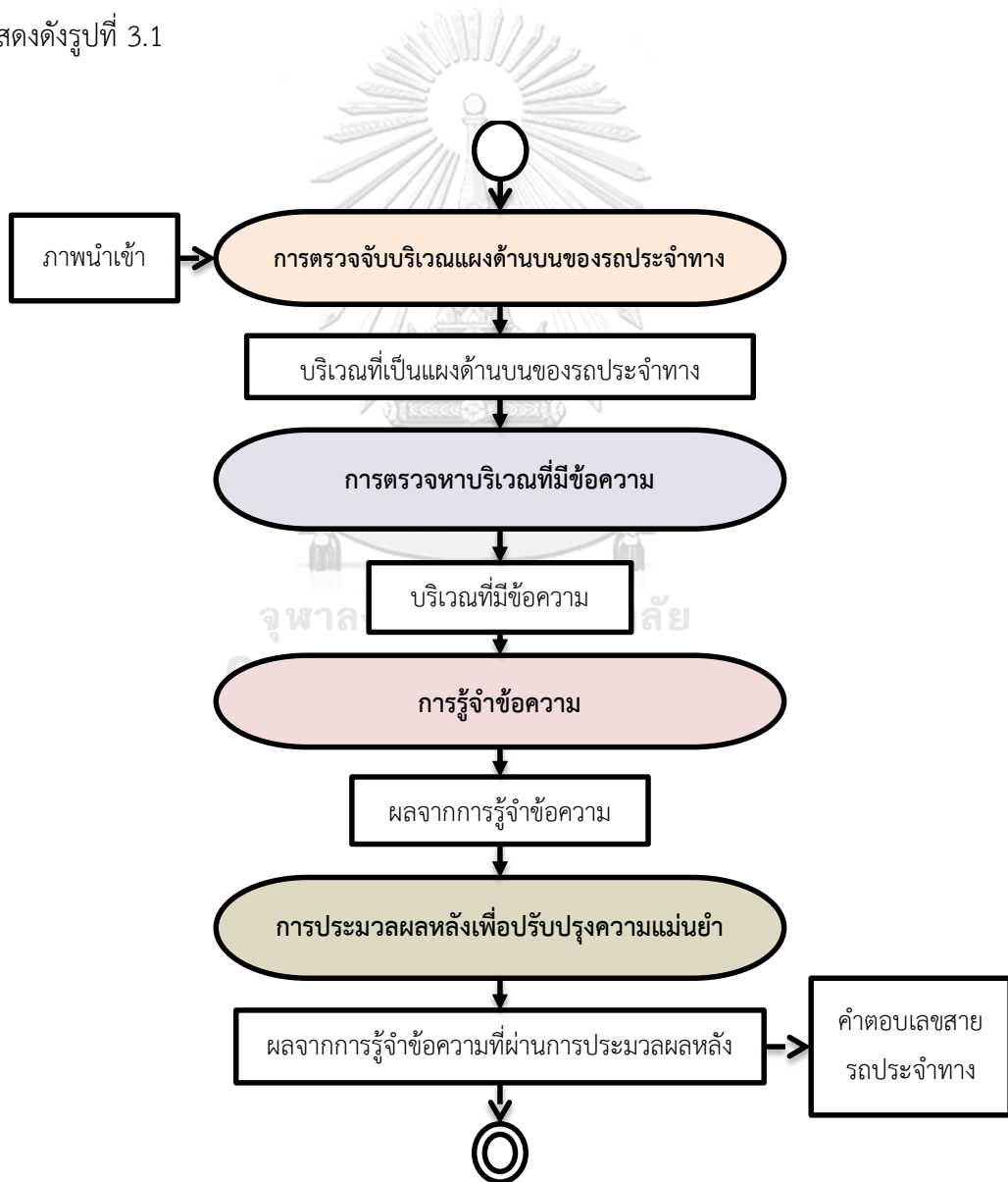
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้ไม่นำรถประจำทางที่แสดงข้อมูลบริเวณแผงด้านบนของรถประจำทางแบบ หลอดไฟแอลอีดีมาใช้ในการทดลอง เนื่องจากรูปถ่ายจากสมาร์ทโฟนไม่สามารถมองเห็นตัวเลขได้ชัดเจน



### บทที่ 3 ขั้นตอนและวิธีที่นำเสนอ

ในบทนี้กล่าวถึงรายละเอียดของขั้นตอนและวิธีที่นำเสนอ ในการตรวจจับหมายเลขสายรถประจำทางจากภาพถ่ายที่ถูกถ่ายด้วยกล้องของสมาร์ทโฟน โดยขั้นตอนวิธีที่นำเสนอ ประกอบไปด้วยขั้นตอนหลัก 4 ขั้นตอน คือ การตรวจจับบริเวณแผงด้านบนของรถประจำทาง การตรวจหาบริเวณที่มีข้อความของสายรถประจำทาง การรู้จำข้อความ และการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ลำดับขั้นตอนโดยภาพรวมของการทำงานของระบบการอ่านเลขสายรถประจำทางจากภาพแสดงดังรูปที่ 3.1



รูปที่ 3.1 แผนผังการทำงานของระบบการอ่านเลขสายรถประจำทางจากภาพ

ภาพที่ใช้ในงานวิจัยนี้ เป็นภาพถ่ายรถประจำทางขององค์การขนส่งมวลชนกรุงเทพ (ขสมก.) หรือรถร่วมบริการ ที่วิ่งอยู่ในกรุงเทพมหานครและปริมณฑล ซึ่งถูกถ่ายออกจากบริเวณป้ายรถประจำทาง เนื่องจากในภาพประกอบไปด้วยสิ่งต่างๆ เช่น รถยนต์ บุคคล ป้ายโฆษณา ฯลฯ นอกเหนือจากรถประจำทาง ซึ่งแสดงดังรูปที่ 3.2 ดังนั้นจึงต้องมีกระบวนการหาตำแหน่งของบริเวณแผงด้านบนรถประจำทางก่อน เพื่อจะได้ตัดส่วนภาพอื่นๆ ที่ไม่เกี่ยวข้องในการอ่านเลขสายรถประจำทางออกไป เป็นการลดบริเวณที่จะค้นหา (Search Space) และเพิ่มประสิทธิภาพในการสกัดและอ่านเลขสายรถประจำทาง



(ก)



(ข)



(ค)



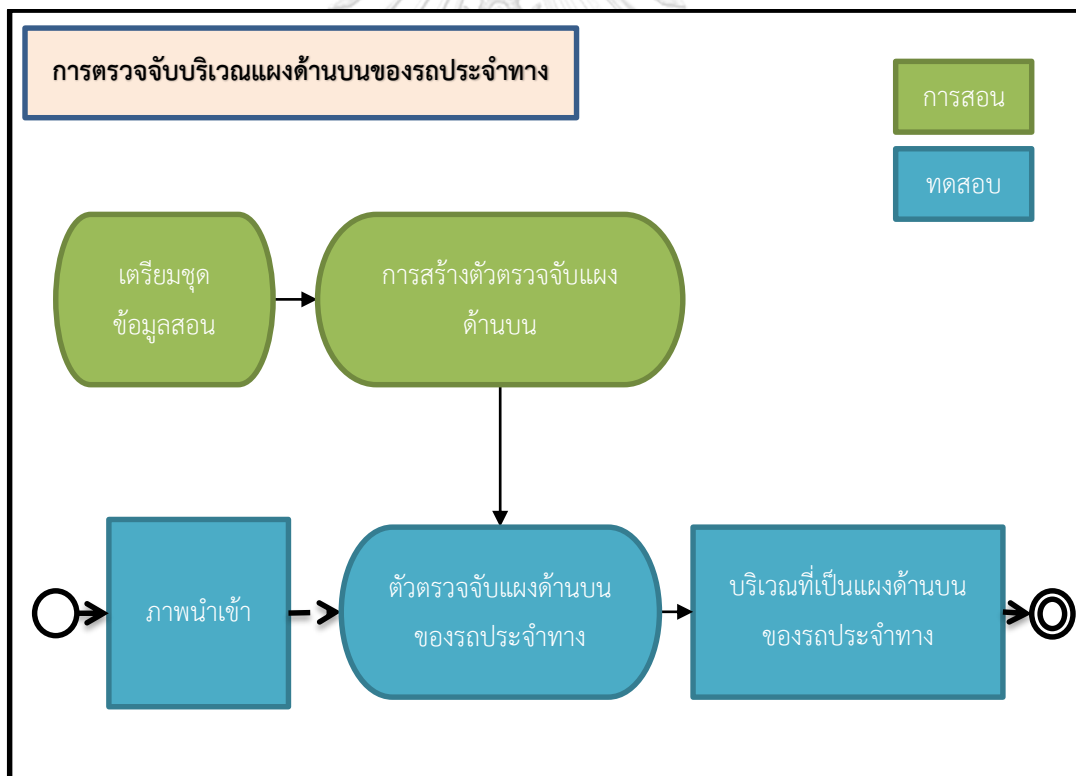
(ง)

รูปที่ 3.2 ตัวอย่างภาพถ่ายรถประจำทาง ขสมก. หรือรถร่วมบริการ ที่ใช้ในวิทยานิพนธ์ฉบับนี้ ซึ่งมีความหลากหลายและซับซ้อนของทั้งฉากหลัง รวมถึงการบดบังจากอ็อบเจกต์ต่าง ๆ (ก)(ข)(ค)(ง) แสดงภาพถ่ายรถประจำทางที่ในภาพประกอบไปด้วยสิ่งทีนอกเหนือจากรถประจำทาง เช่น รถยนต์ บุคคล หรือป้ายโฆษณา

### 3.1 ขั้นตอนการตรวจจับสนบริเวณแฉงด้านบนของรถประจําทาง

ในส่วของการตรวจจับสนบริเวณแฉงด้านบนของรถประจําทางจากภาพนำเข้า เนื่องจากความหลากหลายของภาพนำเข้า เช่น สภาพแสง ภาพเบลอ มุมกล้อง รวมถึงลักษณะรูปร่าง ขนาด สี ของแฉงด้านบนของรถประจําทางที่มีความหลากหลายไม่เหมือนกัน ทำให้การใช้วิธีการทางด้านการวิเคราะห์จากลักษณะเด่นในภาพ เป็นเรื่องยากในการตรวจจับสนให้ถูกต้องและแม่นยำ ดังนั้นในวิทยานิพนธ์ฉบับนี้ จึงได้เลือกใช้ตัวตรวจจับสนที่สามารถสร้างจากวิธีการการเรียนรู้ของเครื่องที่สามารถรู้คุณลักษณะสำคัญจากชุดข้อมูลที่สอน โดยวิธีการคอนโวลูชันนอลนิรอรลเน็ตเวิร์คแบบพาสเตอร์อาร์ซีเอ็นเอ็น ที่ได้อธิบายขั้นตอนการทำงานในหัวข้อที่ 2.1.2 และต้องมีการเตรียมชุดข้อมูลสอนเพื่อสร้างตัวตรวจจับสนบริเวณแฉงด้านบนของรถประจําทาง

การตรวจจับสนบริเวณแฉงด้านบนของรถประจําทาง แบ่งออกเป็น 2 ขั้นตอน คือ ขั้นตอนการสอน และขั้นตอนการทดสอบ แสดงดังรูปที่ 3.3



รูปที่ 3.3 ขั้นตอนของการตรวจจับสนบริเวณแฉงด้านบนของรถประจําทาง

### 3.1.1 ขั้นตอนการสอน

#### 3.1.1.1 การเตรียมชุดข้อมูลสอน

ชุดข้อมูลสอนเป็นภาพถ่ายรถประจำทางที่ถูกถ่ายจากสมาร์ทโฟน โดยทำการสุ่มเลือกภาพถ่าย จำนวน 80% เป็นชุดข้อมูลสอน และชุดข้อมูลทดสอบเป็นชุดข้อมูลที่เหลือ 20% จากภาพถ่ายของรถประจำทางทั้งหมด ที่มีผู้ถ่ายภาพทั้งหมด 5 คน เพื่อความหลากหลายของมุมมองภาพรถประจำทางที่ถูกถ่าย หลังจากนั้นย่อขนาดของภาพลง 4 เท่าจากขนาดภาพดั้งเดิม เนื่องจากหากใช้ภาพขนาดดั้งเดิมในการทำการสอน จะทำให้เมมโมรี่ไม่เพียงพอต่อการทำการสอน จากนั้นทำการหาพิกัด 4 จุด ที่ล้อมรอบบริเวณที่เป็นแผงด้านบนของรถประจำทาง โดยใช้โปรแกรม labellmg [37] ในการตีกรอบบริเวณที่เป็นแผงด้านบนของรถประจำทาง โปรแกรมจะเก็บข้อมูลพิกัดจุด  $X_{min}$   $Y_{min}$   $X_{max}$   $Y_{max}$  จากภาพรถประจำทาง และเก็บอยู่ในรูปแบบปาสคาล วีไอซี (PASCAL VOC) [25] ซึ่งเป็นไฟล์เอกซ์เอ็มแอล (XML) จากนั้นทำการแปลงไฟล์เอกซ์เอ็มแอล เป็นไฟล์ ซีเอสวี (CSV) เพื่อเตรียมนำไปใช้ในการสอนข้อมูล โดยรูปที่ 3.4 แสดงถึงภาพของโปรแกรม labellmg ที่ใช้ในการเก็บพิกัดข้อมูล



รูปที่ 3.4 โปรแกรม labellmg ที่ใช้ในการเก็บพิกัดข้อมูล

ในงานวิจัยนี้ได้เลือกใช้เทนเซอร์ฟล [21] ซึ่งเป็นเครื่องมือและไลบรารีช่วยในการทำการเรียนรู้เชิงลึก (Deep Learning) มาช่วยในการทำการสอนข้อมูล ซึ่งต้องจัดเตรียมข้อมูลให้อยู่ในรูปของไฟล์ที่เอฟเรคคอดส์ (TFRecords) โดยต้องทำการแปลงไฟล์ซีเอสวี ที่เก็บข้อมูลพิกัดที่เป็น

แผงด้านบนของรถประจำทาง ทั้งชุดข้อมูลสอน และชุดข้อมูลทดสอบให้อยู่ในรูปของไฟล์ที่เอพเรคคอดส์โดยใช้โปรแกรมจาก [38] เพื่อนำมาใช้เป็นข้อมูลในการทำการสอน

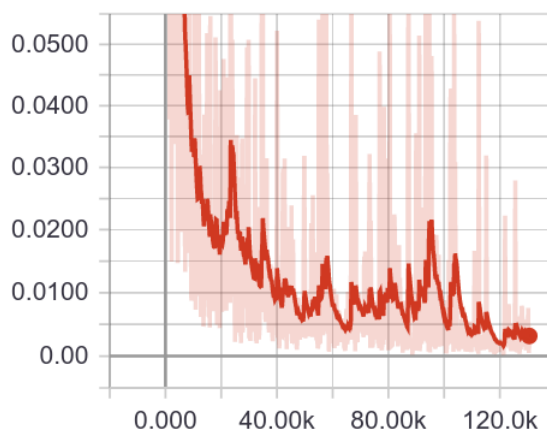
### 3.1.1.2 การสร้างตัวตรวจจับแฉงด้านบนของรถประจำทาง

ในกระบวนการสอนข้อมูลต้องใช้ตัวแบบการถ่ายโอนการเรียนรู้ ซึ่งผู้วิจัยได้เลือกใช้ตัวแบบการถ่ายโอนการเรียนรู้แบบ เรสเน็ต101 (ResNet101) [15] ที่ถูกสอนมาจากคลังภาพของ ไมโครซอฟท์ โคโค (Microsoft COCO) [13] ด้วยวิธีการฟาสเตอร์อาร์ซีเอ็นเอ็น ซึ่งเป็นวิธีการที่ได้รับความนิยมในการนำมาตรวจจับวัตถุ และให้ผลที่แม่นยำ

กระบวนการสอนต้องมีการตั้งค่าจากไฟล์คอนฟิก (Config File) ที่ชื่อไฟล์เป็นชื่อของตัวแบบการถ่ายโอนการเรียนรู้และวิธีการที่ใช้ เพื่อกำหนดค่าจำนวนของคลาสที่สอน เส้นทางที่อยู่ของไฟล์ที่เตรียมไว้สำหรับชุดข้อมูลสอน และเส้นทางที่อยู่ของตัวแบบการถ่ายโอนการเรียนรู้ที่ใช้ หลังจากนั้นจึงเริ่มทำการสอน

ขณะทำการสอน ค่าความเสียหายทั้งหมด (Total Loss) ซึ่งเป็นค่าที่บอกถึงการทำนายของตัวแบบ ว่ามีความใกล้เคียงกับค่าความจริงมากน้อยแค่ไหน หากค่าความเสียหายมีค่าน้อย ตัวแบบจะมีประสิทธิภาพที่ดีขึ้น ดังรูปที่ 3.5 แสดงผลของค่าความเสียหายทั้งหมดหลังจากการสอน

TotalLoss



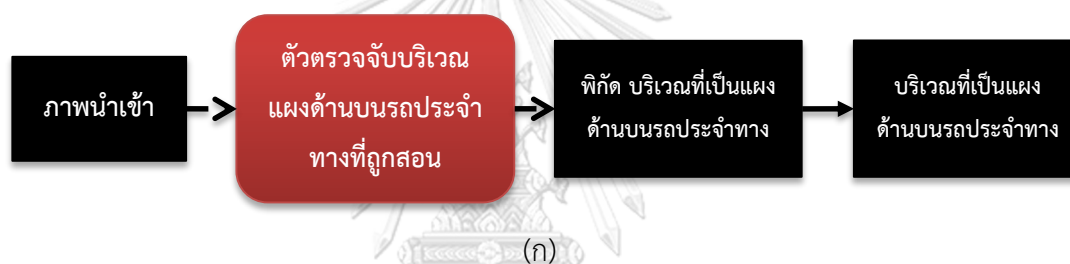
รูปที่ 3.5 กราฟค่าความเสียหายทั้งหมด ของการสอน

หลังจากเสร็จสิ้นการทำการสอน ผู้วิจัยได้เลือกใช้ตัวแบบที่ถูกสอนในการวนซ้ำ (Iteration) ที่ 130,563 เนื่องจากค่าความเสียหายทั้งหมด มีแนวโน้มเริ่มเข้าใกล้ค่า 0 อย่างต่อเนื่อง และผู้วิจัยเห็นว่าค่าความเสียหายทั้งหมดมีการเปลี่ยนแปลงไม่มากนักในช่วงระยะเวลาของการวนซ้ำที่ 120,000 เป็นต้นไป จึงทำการหยุดสอนแบบกำหนดเอง (Manual) จากนั้นทำการส่งออกตัวแบบที่ถูกสอน

หลังจากการวนซ้ำที่ 130,563 เพื่อนำมาใช้เป็นตัวอย่างในการตรวจจับบริเวณแฉงด้านบนของรถประจำทาง

### 3.1.2 ขั้นตอนการทดสอบ

การทดสอบตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง เริ่มจากภาพนำเข้าที่เป็นภาพถ่ายรถประจำทาง นำมาเข้าสู่ตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางที่ทำการสอน จากนั้นตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางจะตอบคำตบออกมาเป็นเซตของจุดพิกัด 4 จุด ของบริเวณที่เป็นแฉงด้านบนของรถประจำทาง โดยขั้นตอนการทดสอบและตัวอย่างรูปภาพที่ทำการเข้าตัวแบบตรวจจับบริเวณแฉงด้านบนของรถประจำทาง แสดงดังรูปที่ 3.6 และรูปที่ได้หลังจากการตรวจจับแฉงด้านบนของรถประจำทางแสดงดังรูปที่ 3.7



(ข)

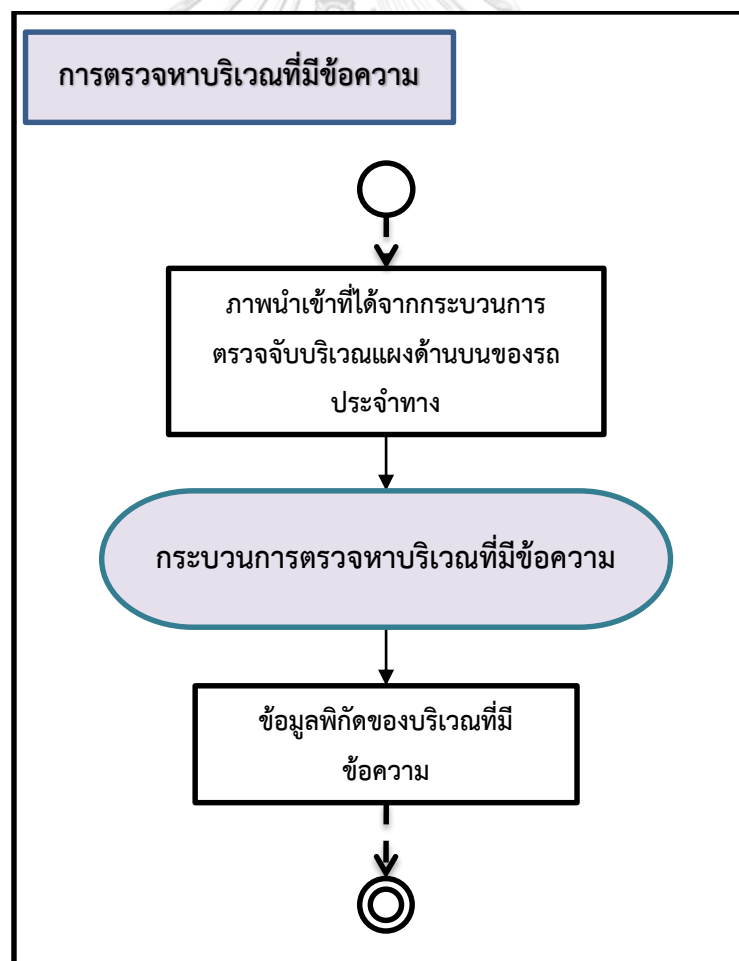
รูปที่ 3.6 การทดสอบตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง (ก) ขั้นตอนการทดสอบ (ข) ตัวอย่างรูปภาพหลังจากเข้าตัวตรวจจับแฉงด้านบนของรถประจำทาง โดยกรอบสี่เหลี่ยมสีเขียวแสดงถึงบริเวณแฉงด้านบนของรถประจำทาง



รูปที่ 3.7 รูปที่ได้หลังจากการตรวจจับบริเวณแผงด้านบนของรถประจำทาง

### 3.2 การตรวจหาบริเวณที่มีข้อความ

หลังจากได้บริเวณแผงด้านบนของรถประจำทาง ขั้นตอนต่อไปคือนำภาพบริเวณแผงด้านบนของรถประจำทาง ส่งเข้าสู่กระบวนการตรวจจับบริเวณที่มีข้อความ โดยผู้วิจัยได้เลือกใช้ตัวตรวจจับบริเวณที่มีข้อความโดยใช้การเรียนรู้เชิงลึกด้วยวิธีฟาสเตอร์อาร์ซีเอ็นเอ็นจากธนานพ และคณะ [39], [40] โดยรูปที่ 3.8 แสดงการทำงานของการทำงานของการตรวจหาบริเวณที่มีข้อความ



รูปที่ 3.8 แสดงขั้นตอนการทำงานของการทำงานของการตรวจหาบริเวณที่มีข้อความ

### 3.2.1 กระบวนการตรวจหาบริเวณที่มีข้อความ

ตัวตรวจจับข้อความใช้กระบวนการคอนโวลูชันนอลนิวรัลเน็ตเวิร์ค แบบฟาสเตอร์อาร์ซีเอ็นเอ็น เช่นเดียวกับ ตัวตรวจจับแผงด้านบนของรถประจำทาง โดยตัวตรวจจับข้อความนี้ถูกสอน จากชุดข้อมูลภาพตัวอักษรที่มีทั้ง ภาษาไทย ภาษาอังกฤษ และตัวเลข จากฐานข้อมูลอิมเมจเน็ต (ImageNet) [12] และฐานข้อมูลของไอซีดีเออาร์ (ICDAR) [41] โดยใช้ตัวแบบการถ่ายโอนการเรียนรู้ที่ถูกสอนจากคลังภาพของฐานข้อมูล อิมเมจเน็ต คำตอบที่ได้จากกระบวนการตรวจหาบริเวณที่มีข้อความ ตบออกมาเป็นข้อมูลพิกัด 4 ตำแหน่ง ที่ล้อมรอบบริเวณที่มีข้อความ โดยภาพตัวอย่างของการหาบริเวณที่มีข้อความ แสดงดังรูปที่ 3.9

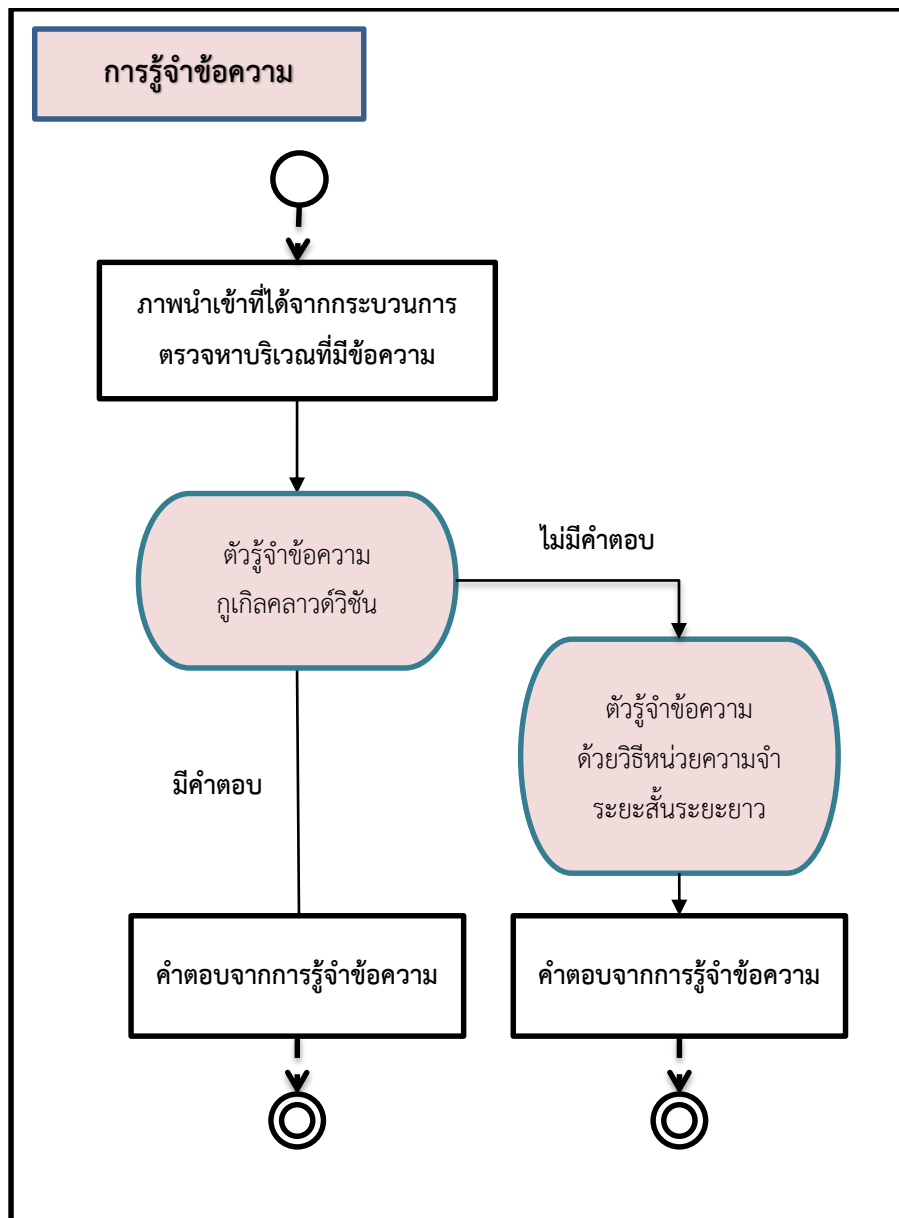


**รูปที่ 3.9** ตัวอย่างผลลัพธ์ที่ได้จากการตรวจหาบริเวณที่มีข้อความ (ก)(ข) ตัวอย่างที่ได้จากการตรวจหาบริเวณที่มีข้อความที่บริเวณแผงด้านบนรถประจำทางแสดงเฉพาะเลขสายรถประจำทาง (ค) ตัวอย่างที่ได้จากการตรวจหาบริเวณที่มีข้อความที่บริเวณแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทาง และชื่อสถานที่



### 3.3 การรู้จำข้อความ

เมื่อสามารถหาบริเวณที่มีข้อความจากขั้นตอนการตรวจหาบริเวณที่เป็นข้อความได้แล้ว ขั้นตอนต่อไปเป็นการนำบริเวณภาพที่ได้จากการตรวจหาบริเวณที่มีข้อความ มาเข้าสู่ตัวรู้จำข้อความ ซึ่งตัวรู้จำข้อความจะทำการแปลงออกมาเป็นตัวอักษร โดยงานวิจัยนี้ได้เลือกใช้ตัวรู้จำข้อความจำนวน 2 ตัว คือตัวรู้จำข้อความจากกูเกิลคลาวด์วิชัน (Google Cloud Vision) [42] และหากคำตอบที่ได้ไม่มีคำตอบจากตัวรู้จำกูเกิลคลาวด์วิชัน จะนำบริเวณภาพนั้นเข้าสู่ตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาวของธนานพ และคณะ [39], [40] โดยขั้นตอนการทำงานของตัวรู้จำข้อความ แสดงดังรูปที่ 3.10



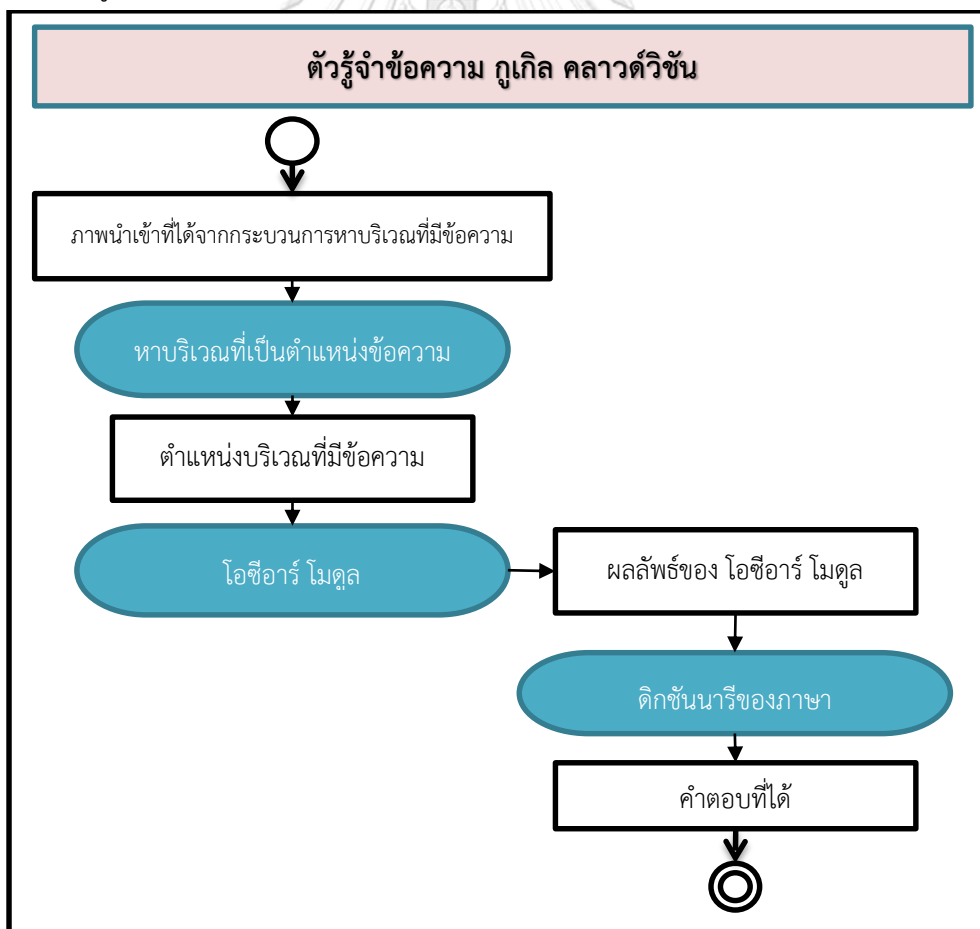
รูปที่ 3.10 ขั้นตอนการทำงานของตัวรู้จำข้อความ

### 3.3.1 ตัวรู้จำข้อความ กูเกิลคลาวด์วิชัน (Google Cloud Vision Text Recognition)

[43]

ตัวรู้จำข้อความจากกูเกิลคลาวด์วิชัน เป็นส่วนหนึ่งของ กูเกิลคลาวด์วิชัน เอพีไอ (Google Cloud Vision API) ที่เปิดให้นักพัฒนาสามารถเข้าไปใช้ระบบของกูเกิล ที่พัฒนาเกี่ยวกับทางด้าน การประมวลผลภาพ โดยเฉพาะตัวรู้จำข้อความที่เป็นที่นิยมอย่างมาก เนื่องจากทรัพยากรและข้อมูลของ กูเกิลที่มีจำนวนมาก จึงสามารถทำให้รู้จำตัวอักษรได้อย่างแม่นยำ และยังสามารถรู้จำได้หลากหลาย ภาษา

การทำงานของตัวรู้จำข้อความกูเกิลคลาวด์วิชัน เริ่มจากหาบริเวณตำแหน่งที่มีข้อความจาก ภาพนำเข้า หลังจากที่ได้ทราบตำแหน่งข้อความ โอซีอาร์โมดูล (OCR Module) ที่เอาไว้ใช้รู้จำตัวอักษร ถูกนำมาใช้วิเคราะห์บริเวณตำแหน่งที่มีข้อความ หลังจากนั้นข้อความที่ถูกรู้จำออกมา ถูกนำไปสู่ กระบวนการประมวลผลหลัง ซึ่งนำไปเทียบว่าข้อความที่รู้จำมานั้น มาจากภาษาใด และส่งไป เปรียบเทียบกับดิกชันนารีของภาษานั้น โดยขั้นตอนการทำงานของตัวรู้จำข้อความ กูเกิลคลาวด์ วิชัน แสดงดังรูปที่ 3.11



รูปที่ 3.11 ขั้นตอนการทำงานของตัวรู้จำข้อความจากกูเกิลคลาวด์วิชัน

การใช้ตัวรู้จำข้อความของกูเกิล คลาวด์วิชันได้ ต้องสมัครใช้บริการของ กูเกิล ดีเวลอปเปอร์ (Google Developers) [44] และเข้าไปในส่วนของ กูเกิล คลาวด์วิชัน เอพีไอ เพื่อขอรหัส (Key) ถึงสามารถนำมาใช้บริการเกี่ยวกับกูเกิลคลาวด์วิชันได้ และการขอใช้ตัวรู้จำข้อความของ กูเกิล คลาวด์วิชัน ต้องทำการเขียนโปรแกรมติดต่อกับกูเกิลคลาวด์วิชันเพื่อขอเรียกใช้งาน โดยมีตัวอย่างโปรแกรมที่ใช้ในการเขียนติดต่อกับเว็บไซต์ของกูเกิลคลาวด์วิชัน ผลลัพธ์ของคำตอบที่ตอบออกมา อยู่ในรูปแบบของเจสัน (JSON Format) ที่มีคำตอบหลากหลายหมวด เช่น ตัวอักษรที่ตรวจจับได้ (description) ภาษา (languageCode) หรือ ตำแหน่งที่มีตัวอักษรอยู่ (boundingPoly) โดยงานวิจัยนี้ เลือกเฉพาะตัวอักษรที่ตรวจจับได้ โดยตัวอย่างของคำตอบที่ตอบออกมาจากกูเกิลคลาวด์วิชัน แสดงดังรูปที่ 3.12



รูปที่ 3.12 ตัวอย่างคำตอบที่ได้จากกูเกิลคลาวด์วิชัน

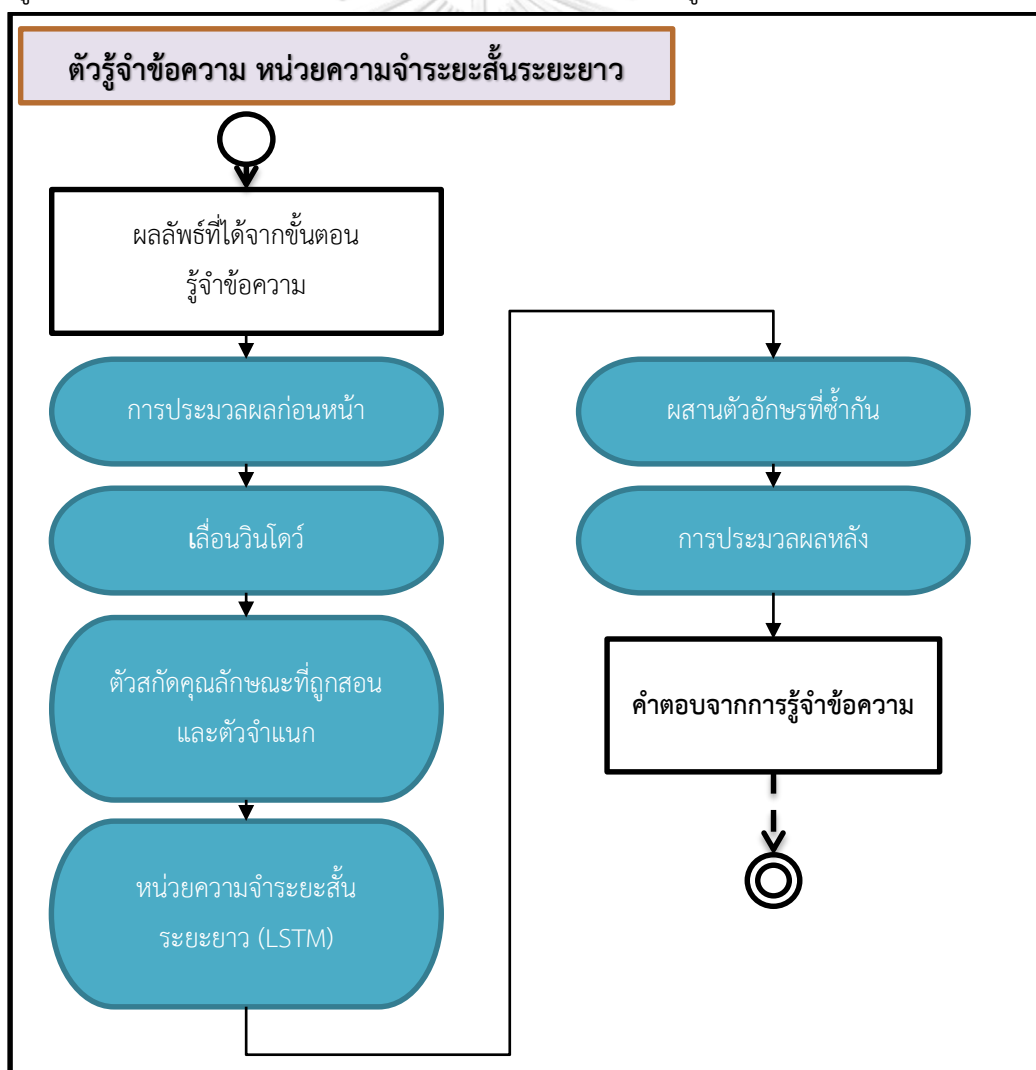
หากคำตอบในหมวดของตัวอักษรที่ตรวจจับได้ มีข้อมูล เมื่อใช้ตัวรู้จำข้อความจากกูเกิล คลาวด์วิชัน จะถูกส่งต่อไปยังขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ แต่หากคำตอบในหมวดของตัวอักษรที่ตรวจจับได้ ไม่มีข้อมูลหรือไม่มีคำตอบ บริเวณรูปนำเข้านั้นจะถูกนำไปเข้ากระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว (แอลเอสทีเอ็ม) เพื่อหาคำตอบต่อไป

### 3.3.2 ตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว [39], [40]

ตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว เป็นตัวรู้จำที่ใช้การเรียนรู้เชิงลึก คอนโวลูชันนอลนิวรอลเน็ตเวิร์ค ในการสอนตัวอักษร จากชุดฐานข้อมูลอิมเมจเน็ต (ImageNet) [12]

และ ไอซีดีเออาร์ (ICDAR) [41] โดยภาพข้อมูลของตัวอักษร ประกอบด้วยตัวอักษรภาษาอังกฤษ และ ตัวเลข

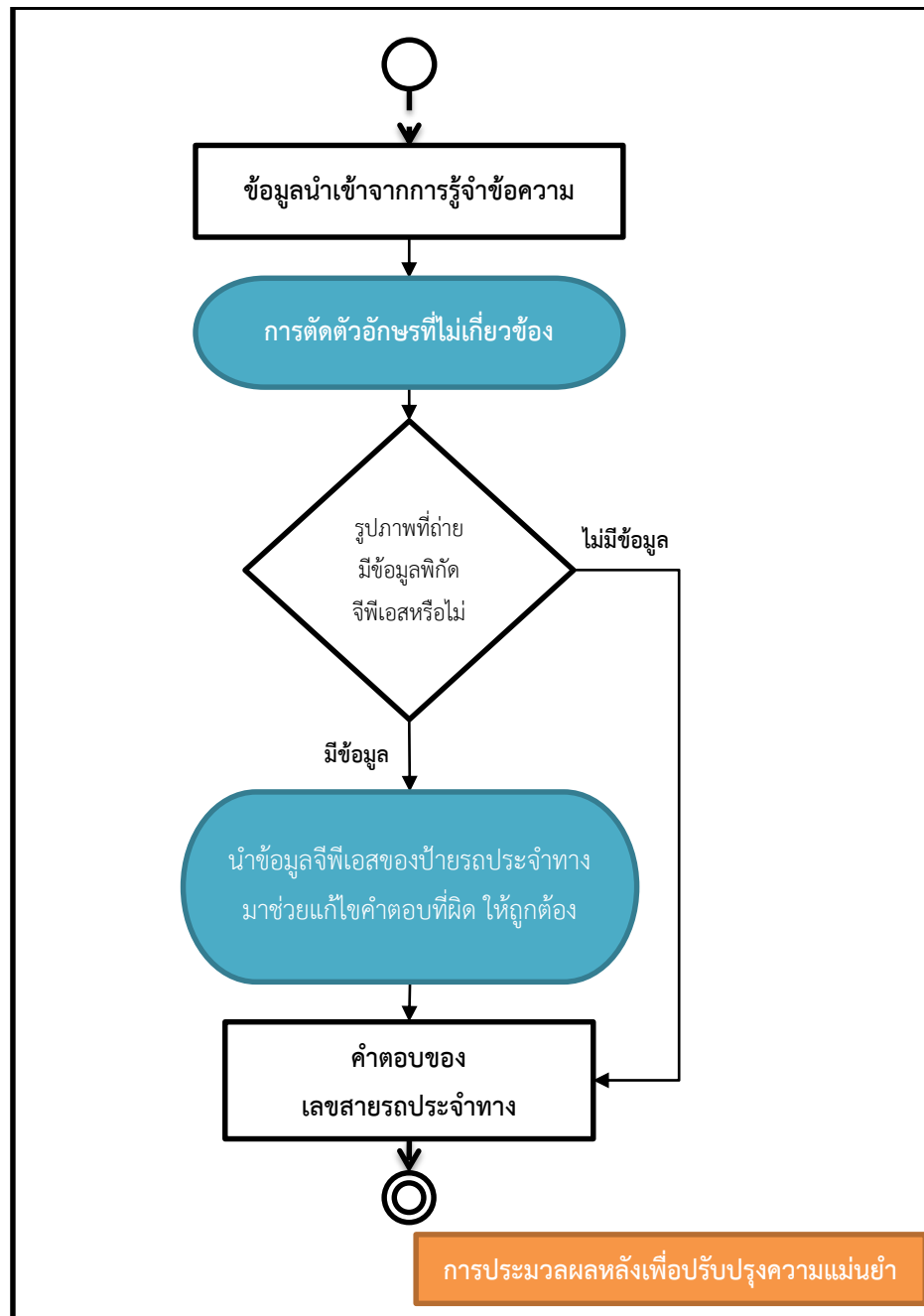
กระบวนการการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว เริ่มต้นจากบริเวณภาพ นำเข้าที่ไม่สามารถตรวจจับข้อความได้จากตัวรู้จำข้อความภูเกิล คลาวด์วิชัน นำเข้ากระบวนการประมวลผลก่อนหน้าซึ่งจะทำการนอมอลไลเซชัน ให้ภาพมีความสูงขนาด 32 พิกเซล และทำการเลื่อนวินโดว์ด้วยขนาด 32x32 พิกเซล จากซ้ายไปขวา จากนั้นส่งไปยังตัวสกัดคุณลักษณะที่ถูกลองจนจำแนกออกมาเป็นตัวอักษร ด้วยวิธีการคอนโวลูชันนอลนิวโรลเน็ตเวิร์ค และนำเข้าสู่กระบวนการหน่วยความจำระยะสั้นระยะยาวเพื่อทำการจดจำ เมื่อทำงานเลื่อนวินโดว์ครบจะได้กลุ่มคำที่มีตัวอักษรซ้ำกันตามแต่ละวินโดว์ที่เลื่อนไป จากนั้นทำการผสานตัวอักษรที่ซ้ำกันให้กลายเป็นตัวอักษรที่น่าจะเป็นเพียงตัวเดียว และเข้าสู่กระบวนการประมวลผลภายหลังจนได้คำตอบ โดยขั้นตอนกระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว แสดงดังรูปที่ 3.13



รูปที่ 3.13 ขั้นตอนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว

เมื่อรวมคำตอบที่ตอบจากตัวรู้จำข้อความทั้งตัวรู้จำข้อความกุเกิลคลาวด์วิชัน และตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว ขั้นตอนต่อไปจะเข้าสู่กระบวนการการประมวลผลหลัง เพื่อปรับปรุงความแม่นยำ

### 3.4 การประมวลผลหลัง เพื่อปรับปรุงความแม่นยำ



รูปที่ 3.14 ลำดับขั้นตอนของการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ

ในกระบวนการนี้ เป็นกระบวนการทำการแก้ไขคำตอบจากส่วนของการรู้จำข้อความที่ตอบผิดให้ถูกต้องมากขึ้น โดยใช้การตัดตัวอักษรที่ไม่เกี่ยวข้องกับข้อมูลของเลขสายรถประจำทาง และใช้ข้อมูลจีพีเอสของภาพถ่ายที่ได้จากการเปิดตำแหน่งโลเคชันเซอร์วิส (Location Service) บนสมาร์ตโฟนขณะทำการถ่ายภาพบริเวณป้ายรถประจำทาง กับข้อมูลจีพีเอสที่เป็นพิกัดตำแหน่งของป้ายรถประจำทาง ในการเปรียบเทียบดูว่า สายรถประจำทางจากภาพที่ถูกถ่ายบริเวณป้ายรถประจำทาง ตรงกับข้อมูลสายรถประจำทางที่ผ่านป้ายรถประจำทางที่มีระยะใกล้ที่สุดหรือไม่ โดยลำดับขั้นตอนของการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ แสดงดังรูปที่ 3.14

ความแม่นยำของข้อมูลจีพีเอสของภาพถ่ายที่ได้จากการเปิดตำแหน่งโลเคชันเซอร์วิสบนสมาร์ตโฟนขณะทำการถ่ายภาพบริเวณป้ายรถประจำทาง จากการทดลองเพื่อหาความแม่นยำของตำแหน่งจีพีเอสบนสมาร์ตโฟนของ van Diggelen และ Enge [45] โดยรวบรวมข้อมูลจากสมาร์ตโฟนมากกว่า 10,000 เครื่อง จาก 100 กว่าประเทศทั่วโลก พบว่าตำแหน่งจีพีเอสบนสมาร์ตโฟนจะมีความคลาดเคลื่อนในระยะรัศมีไม่เกิน 5 เมตร ภายใต้สภาพท้องฟ้าที่เปิดโล่ง ซึ่งจะมีความคลาดเคลื่อนที่รัศมีเพิ่มขึ้นหากบริเวณรอบๆ มีตึกสูง สะพาน ต้นไม้ หรือสิ่งที่ยับบังสัญญาณจากดาวเทียม

#### 3.4.1 การตัดตัวอักษรที่ไม่เกี่ยวข้อง

เนื่องจากขั้นตอนการตรวจหาบริเวณที่มีข้อความจากภาพแนวด้านบนของรถประจำทาง ตรวจจับบริเวณที่มีตัวหนังสือ โดยลักษณะรูปแบบตัวหนังสือบริเวณแนวด้านบนของรถประจำทางมีลักษณะของข้อมูลตัวหนังสืออยู่ 3 รูปแบบตามที่ได้กล่าวเบื้องต้นในหัวข้อที่ 2.2.3 คือ แบบตัวเลขสายรถประจำทางเพียงอย่างเดียว แบบตัวเลขสายรถประจำทางกับข้อมูลสถานที่ที่ผ่าน และแบบตัวเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ โดยแสดงดังรูปที่ 3.15



(ก)



(ข)



(ค)

**รูปที่ 3.15** ลักษณะรูปแบบตัวหนังสือบริเวณแผงด้านบนของรถประจำทาง (ก) แบบตัวเลขสายรถประจำทางอย่างเดียว (ข) แบบตัวเลขสายรถประจำทางกับข้อมูลสถานที่ที่ผ่าน (ค) แบบตัวเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ

1) แบบตัวเลขสายรถประจำทางอย่างเดียว คือ บริเวณแผงด้านบนของรถประจำทางปรากฏเพียงเลขสายรถประจำทาง

2) แบบตัวเลขสายรถประจำทางกับข้อมูลสถานที่ที่ผ่าน คือ บริเวณแผงด้านบนของรถประจำทาง ปรากฏทั้งเลขสายรถประจำทาง และข้อมูลของสถานที่ที่รถประจำทางวิ่ง

3) แบบตัวเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ คือ บริเวณแผงด้านบนของรถประจำทาง ปรากฏเลขสายรถประจำทางและอักษรตัวย่อ เช่น ปอ, ก, A, S ซึ่งเป็นสายที่มีความหมายพิเศษแตกต่างกันไป ตัวอย่างเช่น ปอ หมายถึงรถปรับอากาศ ปอ 14 หมายถึง รถปรับอากาศสาย 14 ซึ่งวิ่งเส้นทางเดียวกันกับสาย 14 อีกตัวอย่างหนึ่งคืออักษร ก ตัวอย่างเช่น สาย 7 กับ 7ก วิ่งในเส้นทางที่แตกต่างกัน ดังนั้น กลุ่มตัวอักษร ปอ, A, S จึงเป็นกลุ่มตัวอักษรที่หากอยู่ด้านหน้า หรือ อักษร ก ต่อท้ายเลขสายรถประจำทาง อาจทำให้มีผลต่อการตัดสินใจในการใช้รถประจำทาง

ดังนั้นผลลัพธ์ที่ได้จากขั้นตอนรู้จำข้อความ จะถูกนำมาตัดตัวอักษรทุกตัว ยกเว้นตัวอักษร 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ป, อ, ก, A, S โดยการเวียนหาตัวอักษรแต่ละตัวอยู่ในเซตคำตอบของตัวอักษรที่ยกเว้นหรือไม่ แต่เนื่องจากคำตอบที่ได้จากขั้นตอนการรู้จำข้อความนั้น มีคำตอบที่รู้จำออกมาได้ใกล้เคียง เช่น ปอ14 รู้จำข้อความได้ บอ14 หรือ A1 รู้จำออกมาเป็น a1 เป็นต้น จึง

จำเป็นต้องมีการแก้ไขคำตอบให้ถูกต้อง ตัวอย่างเช่น หากมีตัวอักษร บอ ให้แก้เป็น ปอ ดังนั้นเซตของตัวอักษรที่ยกเว้นจึงประกอบไปด้วย 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ปอ, บอ, ป, อ, ก, A, a, S, s

การตัดตัวอักษรที่ไม่เกี่ยวข้องเริ่มจากนำผลลัพธ์คำตอบที่ได้จากขั้นตอนการรู้จำข้อความ มาทำการตัดตัวอักษรทุกตัวออก ยกเว้นตัวอักษรที่อยู่ในเซตของตัวอักษรที่ยกเว้น หลังจากนั้นทำการหาตำแหน่งของตัวเลขตัวแรกสุด และตำแหน่งตัวเลขตัวสุดท้ายเพื่อใช้ในการอ้างอิงในการตรวจสอบว่าด้านหน้าของตำแหน่งตัวเลขแรกมีตัวอักษร ปอ, ป, อ, A, a, S, s หรือไม่ หากมีให้คงไว้ และทำการตรวจสอบหลังตำแหน่งอ้างอิงของเลขตัวสุดท้ายว่ามีตัวอักษร ก หรือไม่ หากมีให้คงไว้

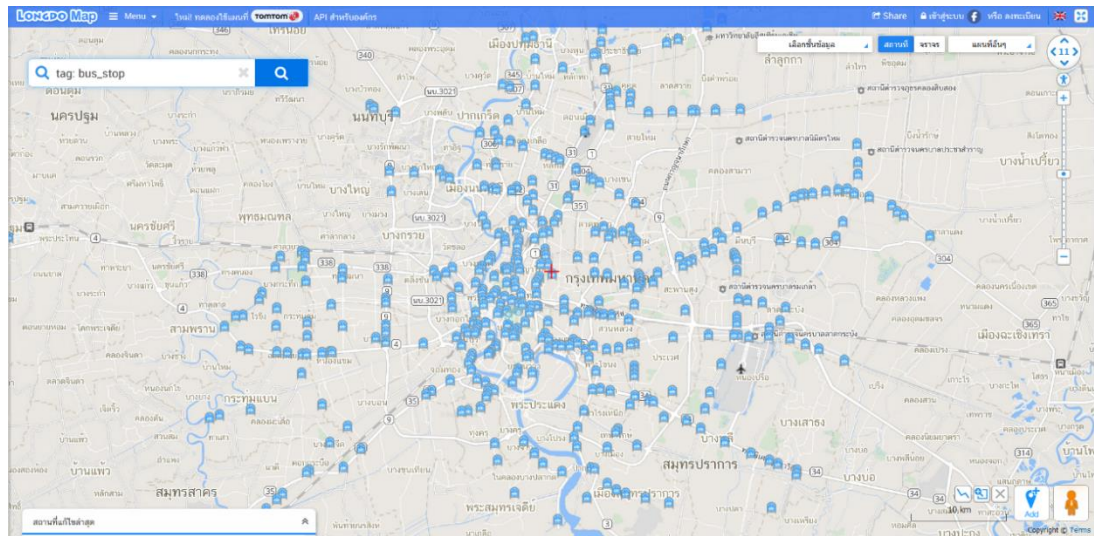
หลังจากแก้ไขคำตอบโดยการตัดตัวอักษรที่ไม่เกี่ยวข้องแล้ว จะนำคำตอบที่ได้เข้าสู่กระบวนการเปรียบเทียบกับข้อมูลจีพีเอสของป้ายรถประจำทาง เพื่อหาคำตอบที่ทำการแก้ไขจากการตัดตัวอักษรที่ไม่เกี่ยวข้อง ใกล้เคียงกับเลขสายรถประจำทางใดที่ผ่านป้ายรถประจำทางนั้น ซึ่งจะดูจากป้ายรถประจำทางที่มีระยะที่ใกล้กับตำแหน่งที่ถ่ายภาพมากที่สุด

ก่อนนำผลลัพธ์ที่ได้จากการตัดตัวอักษรที่ไม่เกี่ยวข้องไปเข้าวิธีการเปรียบเทียบข้อมูลจีพีเอสของป้ายรถประจำทาง จะต้องทำการตรวจสอบว่าภาพถ่ายนำเข้า มีข้อมูลจีพีเอสหรือไม่ หากไม่มีข้อมูลจีพีเอส ให้ตอบผลลัพธ์คำตอบที่ได้จากการตัดตัวอักษรที่ไม่เกี่ยวข้องทันที แต่หากรูปถ่ายที่นำเข้ามามีข้อมูลจีพีเอส ให้เข้าสู่กระบวนการเปรียบเทียบข้อมูลจีพีเอสของป้ายรถประจำทาง

### 3.4.2 ข้อมูลจีพีเอสป้ายรถประจำทาง

การนำข้อมูลจีพีเอสของป้ายรถประจำทางมาช่วยในการค้นหาคำตอบ เป็นการตรวจสอบคำตอบที่ได้จากขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้องว่าคำตอบนั้นถูกต้องหรือไม่ โดยข้อมูลป้ายรถประจำทางที่ได้ใช้ในงานวิจัยนี้เป็นข้อมูลป้ายรถประจำทางจากแผนที่ลองดู (Longdo Maps) [46] ซึ่งเปิดเป็นข้อมูลโอเพนซอส ให้นักพัฒนาสามารถนำข้อมูลไปใช้ในการพัฒนาได้ โดยข้อมูลมีมากกว่า 5,000 ป้ายรถประจำทาง และ 457 สายรถประจำทาง ของ ขสมก. ในเขตกรุงเทพมหานครและปริมณฑล โดยตัวอย่างข้อมูลป้ายรถประจำทางทั่วกรุงเทพมหานครและปริมณฑล แสดงดังรูปที่ 3.16

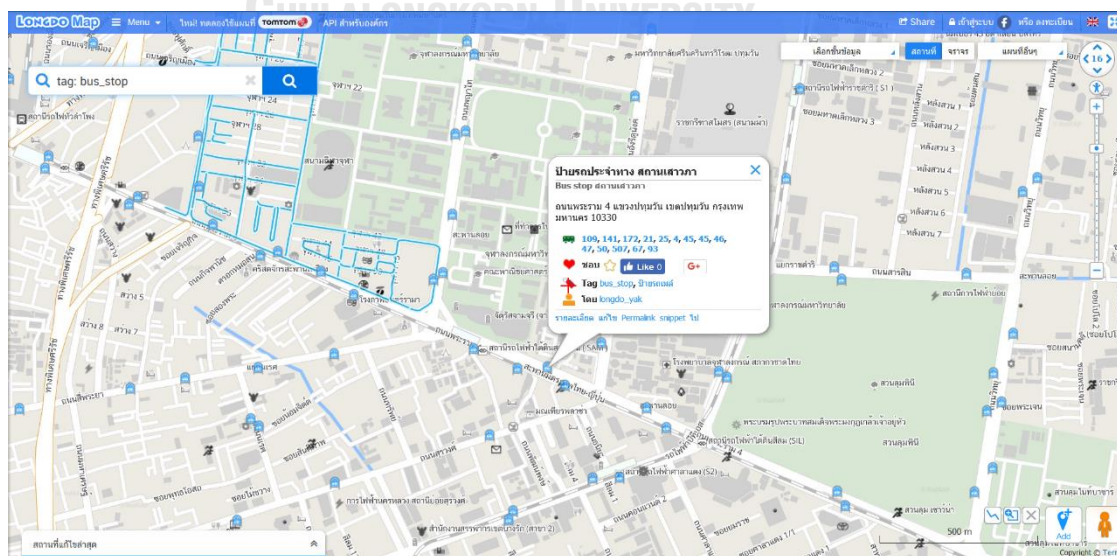




รูปที่ 3.16 ข้อมูลป้ายรถประจำทางทั่วกรุงเทพมหานครและปริมณฑล จากแผนที่ลองดู [47]

โดยแต่ละป้ายรถประจำทางที่ปรากฏในข้อมูลแผนที่ลองดู จะมีข้อมูลบอกว่าป้ายรถประจำทางนี้ มีสายรถประจำทางใดผ่าน ดังตัวอย่างแสดงดังรูปที่ 3.17

การใช้ข้อมูลจีพีเอสของตำแหน่งป้ายสายรถประจำทางมาช่วยในการวิเคราะห์ เริ่มจากนำข้อมูลจีพีเอส พิกัด ละติจูด ลองติจูด ของภาพถ่ายนำเข้ามาเปรียบเทียบกับตำแหน่งข้อมูลจีพีเอส พิกัด ละติจูด ลองติจูด ของป้ายรถประจำทางทั้งหมดเพื่อหาระยะทางที่น้อยที่สุด โดยใช้สูตรฮาเวอไซน์ ที่เป็นการหาระยะทางระหว่างจุดพิกัดจีพีเอส 2 จุด โดยได้อธิบายสูตรฮาเวอไซน์ดังข้อ 2.1.4 หากได้ระยะทางที่น้อยที่สุด ก็ทราบตำแหน่งของป้ายรถประจำทางที่ใกล้ที่สุดว่าเป็นป้ายใด และทราบข้อมูลว่าป้ายรถประจำทางนี้ มีเลขสายรถประจำทางใดผ่าน



รูปที่ 3.17 แผนที่ลองดูที่แสดงข้อมูลป้ายรถประจำทาง ที่มีข้อมูลของสายรถประจำทางที่ผ่าน [47]

เมื่อได้เซตของสายรถประจำทางที่ผ่านป้ายรถประจำทาง จะนำเอาผลลัพธ์ที่ได้จากขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้องมาค้นหาในเซตของสายรถประจำทางที่ผ่านป้ายรถประจำทาง โดยใช้วิธีการระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด ที่ได้อธิบายดังข้อ 2.1.6 ในการหาคำตอบที่ใกล้เคียงระหว่าง ผลลัพธ์คำตอบจากขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้องกับเซตของสายรถประจำทางที่ผ่านป้ายรถประจำทางนั้น วิธีการระยะห่างโดยพิจารณาจากจำนวนการแก้ไขที่น้อยที่สุด จะตอบคำตอบจากเซตของสายรถประจำทางที่ผ่านป้ายรถประจำทางที่มีความน่าจะเป็นที่ใกล้เคียงที่สุด เมื่อเปรียบเทียบกับคำตอบของผลลัพธ์ที่ได้จากการตัดตัวอักษรที่ไม่เกี่ยวข้อง ซึ่งอาจทำให้ในกรณีที่ขั้นตอนการรู้จำข้อความ รู้จำตัวอักษรผิดแต่คำตอบนั้นใกล้เคียงกับคำตอบจริง หรือรู้จำแล้วคำตอบขาดหายไป จะสามารถใช้ข้อมูลสายรถประจำทางที่ผ่านป้ายรถประจำทางนั้น เทียบเคียงเพื่อตอบข้อมูลที่ถูกต้องได้

เนื่องจากข้อมูลป้ายรถประจำทาง และข้อมูลเลขสายรถประจำทางจากแผนที่ลองดูที่ผู้วิจัยนำมาใช้ มีข้อมูลที่ไม่ครบถ้วนและไม่มีข้อมูลประเภทของรถปรับอากาศหรือ ปอ อยู่ในฐานข้อมูลเลขรถประจำทาง จึงมีวิธีการโดยหากได้ผลลัพธ์จากขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้องแล้วมีคำว่า ปอ อยู่ติดกับตัวเลข จะทำการตัด ปอ ออก และนำเฉพาะตัวเลขไปเข้าสู่กระบวนการใช้ข้อมูลจีพีเอสเปรียบเทียบ เพื่อหาเลขสายรถประจำทางที่ต้องการ หลังจากนั้นเมื่อได้คำตอบตัวเลขที่ต้องการแล้ว นำเอา ตัวอักษร ปอ ที่ตัด ใส่มาริเวณด้านหน้าของตัวเลข เพื่อเป็นคำตอบในขั้นสุดท้าย

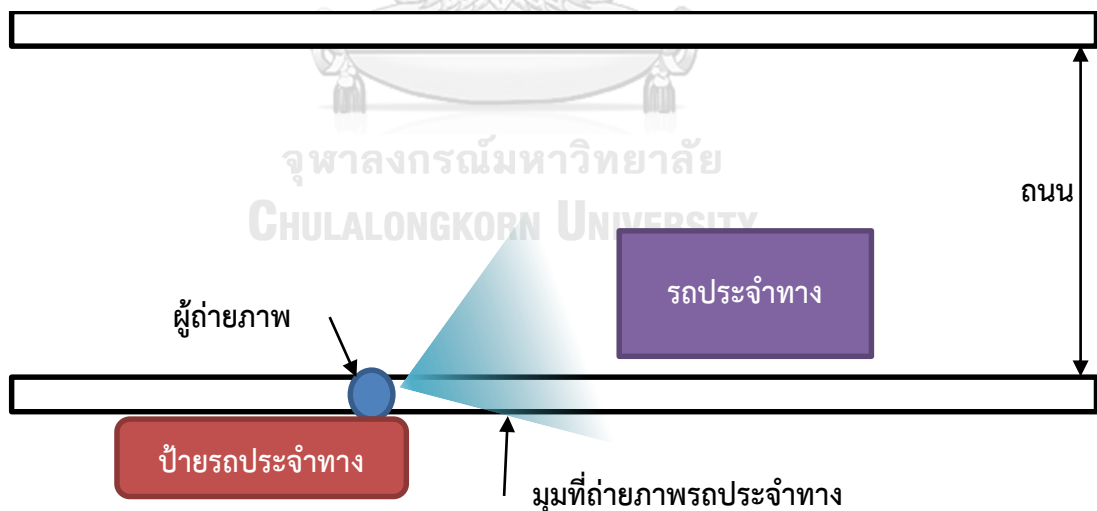
## บทที่ 4

### การทดลองและผลการทดลอง

ในบทนี้กล่าวถึงการทดลองในขั้นตอนตามที่ได้กล่าวมาในบทที่ 3 เพื่อประเมินประสิทธิภาพของระบบอ่านเลขสายรถประจำทางจากภาพ และการวิเคราะห์ผลการทดลอง โดยกล่าวถึง ข้อมูลที่ใช้ในการทดลอง อุปกรณ์ที่ใช้ในการทดลอง การทดลองและผลการทดลองจากขั้นตอนวิธีที่ใช้รวมถึง ปัญหาที่พบ และสรุปผลการทดลอง

#### 4.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลอง เป็นภาพถ่ายรถประจำทางขององค์การขนส่งมวลชนกรุงเทพ (ขสมก.) หรือรถร่วมบริการ ที่ให้บริการอยู่ในกรุงเทพมหานครและปริมณฑล และถ่ายในเวลากลางวัน โดยจุดที่ถ่ายภาพ เป็นจุดที่ยืนอยู่บริเวณป้ายรถประจำทางแล้วถ่ายภาพออกไป ดังรูปที่ 4.1 โดยมีผู้ถ่ายภาพใช้กล้องจากสมาร์ทโฟนของตนเองรวมทั้งหมด 5 คน เพื่อความหลากหลายของมุมมองภาพรถประจำทาง และตั้งค่ากล้องตามค่าปกติ ตารางที่ 4.1 อธิบายสมาร์ทโฟนที่ใช้ถ่ายภาพ และคุณลักษณะของสมาร์ทโฟนที่ใช้ถ่ายภาพ และจำนวนภาพ



รูปที่ 4.1 บริเวณจุดที่ทำการถ่ายภาพรถประจำทาง บริเวณป้ายรถประจำทาง

ตารางที่ 4.1 ชนิดและคุณลักษณะของสมาร์ทโฟนที่ใช้ในการถ่ายภาพ โดยผู้ถ่าย 5 คน

คนที่	สมาร์ทโฟนที่ใช้ถ่าย	คุณลักษณะ	จำนวนภาพ
1	iPhone 5	ความละเอียดกล้อง : 8 ล้านพิกเซล , f/2.4	271
2	iPhone 6	ความละเอียดกล้อง : 8 ล้านพิกเซล , f/2.2	19
3	iPhone 6s	ความละเอียดกล้อง : 12 ล้านพิกเซล , f/2.2	118
4	iPhone 7	ความละเอียดกล้อง : 12 ล้านพิกเซล , f/1.8	105
5	Samsung Galaxy S8	ความละเอียดกล้อง : 12 ล้านพิกเซล , f/1.7	32
<b>รวมจำนวนภาพ</b>			<b>545</b>

#### 4.2 อุปกรณ์ที่ใช้ในการทดลอง

ระบบอ่านเลขสายรถประจำทางจากภาพ ทำงานบนเครื่องคอมพิวเตอร์ที่มีรายละเอียดดังตารางที่ 4.2

ตารางที่ 4.2 รายละเอียดของคอมพิวเตอร์ ที่ใช้ในการประมวลผล

หน่วยประมวลผลหลัก (CPU)	Intel Core i7-4770 3.40 GHz
แรม (RAM)	16 GB
หน่วยประมวลผลกราฟฟิก (GPU)	NVIDIA GeForce GTX 1080 Ti
ระบบปฏิบัติการ (OS)	Linux

#### 4.3 การทดลองและผลการทดลอง

ภาพถ่ายรถประจำทางในการทดลองที่ได้ทำการรวบรวม มีลักษณะของภาพที่แตกต่างกัน เช่น สภาพแสง สภาพแวดล้อม มุมของภาพที่ถูกถ่าย ความซับซ้อนของภาพที่ต่างกัน รวมถึงบางภาพมีรถประจำทางมากกว่า 1 คัน โดยแสดงตัวอย่างภาพถ่ายรถประจำทางที่หลากหลาย ดังรูปที่ 4.2

ภาพถ่ายรถประจำทางที่รวบรวม ถูกแบ่งออกเป็น 2 ส่วน ส่วนแรกนำมาใช้ในการสอนเพื่อหาบริเวณแฉงด้านบนของประจำทาง โดยสุ่มจำนวน 437 ภาพ คิดเป็น 80 เปอร์เซ็นต์ของจำนวนภาพทั้งหมด และส่วนที่สองนำมาใช้ในการทดสอบ โดยสุ่มจำนวน 108 ภาพ คิดเป็น 20 เปอร์เซ็นต์ของจำนวนภาพทั้งหมด



(ก)



(ข)



(ค)

รูปที่ 4.2 ตัวอย่างภาพถ่ายที่ได้ทำการรวบรวม ที่มีลักษณะของภาพที่หลากหลาย (ก) ภาพถ่ายรถประจำทางที่มีผู้คนบนรถประจำทาง (ข) ภาพถ่ายรถประจำทางที่มีเงาสะท้อนบริเวณแผงด้านบนรถประจำทาง (ค) ภาพถ่ายรถประจำทางที่ในภาพมีรถประจำทางมากกว่า 1 คัน

#### 4.3.1 การทดลองตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง

การทดลองแรก เป็นการทดลองตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง โดยเริ่มจากนำภาพนำเข้า เข้าสู่กระบวนการการตรวจจับบริเวณแฉงด้านบนของรถประจำทาง และผลลัพธ์สุดท้ายที่ได้คือบริเวณแฉงด้านบนของรถประจำทาง การประเมินว่าบริเวณแฉงด้านบนของรถประจำทาง ถูกต้องหรือไม่ วัตถุประสงค์จากการประเมินค่าไอโอยู ที่เปรียบเทียบระหว่างพื้นที่ที่ตรวจจับได้ กับพื้นที่ที่เป็นค่าความจริง ซึ่งหากมีค่ามากกว่า 0.7 ถือว่าถูกต้อง

จากภาพนำเข้า ที่นำมาทดสอบจำนวน 108 ภาพ ตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางให้คำตอบที่ตอบออกมาเป็นบริเวณที่เป็นแฉงด้านบนของรถประจำทางทั้งหมด 114 บริเวณ จากค่าความจริงของบริเวณแฉงด้านบนของรถประจำทางที่มีทั้งหมด 113 บริเวณ โดยถูกต้องจำนวน 112 บริเวณ ความผิดพลาดที่เกิดขึ้น ถูกแบ่งออกเป็น 2 กลุ่มคือ ตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางให้คำตอบที่ตอบออกมาเป็นบริเวณแฉงด้านบนของรถประจำทางผิด (ค่าไอโอยูไม่เกิน 0.7) และตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางไม่สามารถตรวจจับบริเวณแฉงด้านบนของรถประจำทางนั้นได้ ซึ่งความผิดพลาดของตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางให้คำตอบที่ตอบออกมาเป็นบริเวณแฉงด้านบนของรถประจำทางผิด (ค่าไอโอยูไม่เกิน 0.7) มีจำนวน 2 ภาพ และ ความผิดพลาดของตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทางไม่สามารถตรวจจับบริเวณแฉงด้านบนของรถประจำทางนั้นได้ มีจำนวน 1 ภาพ โดยผลการทดสอบแสดงดังตารางที่ 4.3

ตารางที่ 4.3 ผลลัพธ์ของการทดลองตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง

ภาพนำเข้า (ภาพ)	จำนวนค่าความจริงของบริเวณ แฉงด้านบนของ รถประจำทาง (บริเวณ)	ผลจากตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง			
		ตอบ ออกมา ทั้งหมด (บริเวณ)	ถูกต้อง (ไอโอยู > 0.7) (บริเวณ) (TP)	ผิด (ไอโอยู < 0.7) (บริเวณ) (FP)	ผิด (ตรวจหาไม่เจอ) (บริเวณ) (FN)
108	113	114	112	2	1

และผลลัพธ์ที่ได้จากตัวตรวจจับบริเวณแฉงด้านบนของรถประจำทาง ถูกประเมินประสิทธิภาพออกมาในค่าความเที่ยง (Precision) ค่ารีคอล (Recall) และค่าการวัดเอฟ (F-Measure) ดังตารางที่ 4.4

ตารางที่ 4.4 การประเมินประสิทธิภาพของตัวตรวจจับบริเวณแฉ่งด้านบนของรถประจำทาง

ตัวตรวจจับบริเวณแฉ่งด้านบนของรถประจำทาง		
ค่าความเที่ยง	ค่ารีคอล	ค่าการวัดเอฟ
0.982	0.991	0.986

จากผลการทำนายว่าเป็นบริเวณแฉ่งด้านบนของรถประจำทาง มีความถูกต้องที่ทำนายว่าเป็นบริเวณแฉ่งด้านบนของรถประจำทาง หรือค่าความเที่ยง เท่ากับ 0.982 และจากจำนวนค่าความจริงของบริเวณแฉ่งด้านบนรถประจำทางทั้งหมด มีความถูกต้องที่ทำนายว่าเป็นบริเวณแฉ่งด้านบนของรถประจำทาง หรือรีคอล เท่ากับ 0.991 และค่าการวัดเอฟ ซึ่งเป็นการวัดประสิทธิภาพโดยรวมของทั้งค่าความเที่ยง และค่ารีคอล มีค่าเท่ากับ 0.986



(ก)



(ข)



(ค)



(ง)

รูปที่ 4.3 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแฉ่งด้านบนของรถประจำทาง (ก)-(ง) ผลลัพธ์ที่ถูกต้อง (TP) ที่ได้จากการทดลองของตัวตรวจจับแฉ่งด้านบนของรถประจำทาง



(ก)



(ข)



(ค)



(ง)

รูปที่ 4.4 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง (ก)-(ง) ผลลัพธ์ที่ถูกต้อง (TP) ที่ได้จากการทดลองของตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง

รูปที่ 4.3 และรูปที่ 4.4 แสดงภาพตัวอย่างจากผลการทดลองของตัวตรวจจับบริเวณแผงด้านบนของรถประจำทางที่ถูกต้อง และนำไปเข้ากระบวนการหาบริเวณที่เป็นข้อความเพื่อวิเคราะห์หาบริเวณที่มีข้อความต่อไป

ในส่วนผลการตรวจจับที่ผิดพลาดมีอยู่ 2 กรณี ในกรณีแรกเนื่องจากตัวตรวจจับบริเวณแผงด้านบนของรถประจำทางคิดว่าบริเวณนั้นเป็นส่วนหนึ่งของบริเวณแผงด้านบนของรถประจำทางจึงทำนายคำตอบออกมา และกรณีที่สองตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง ตรวจจับไม่พบบริเวณแผงด้านบนของรถประจำทาง โดยตัวอย่างผลการตรวจจับที่ผิดพลาด แสดงดังรูปที่ 4.5





(ก)



(ข)

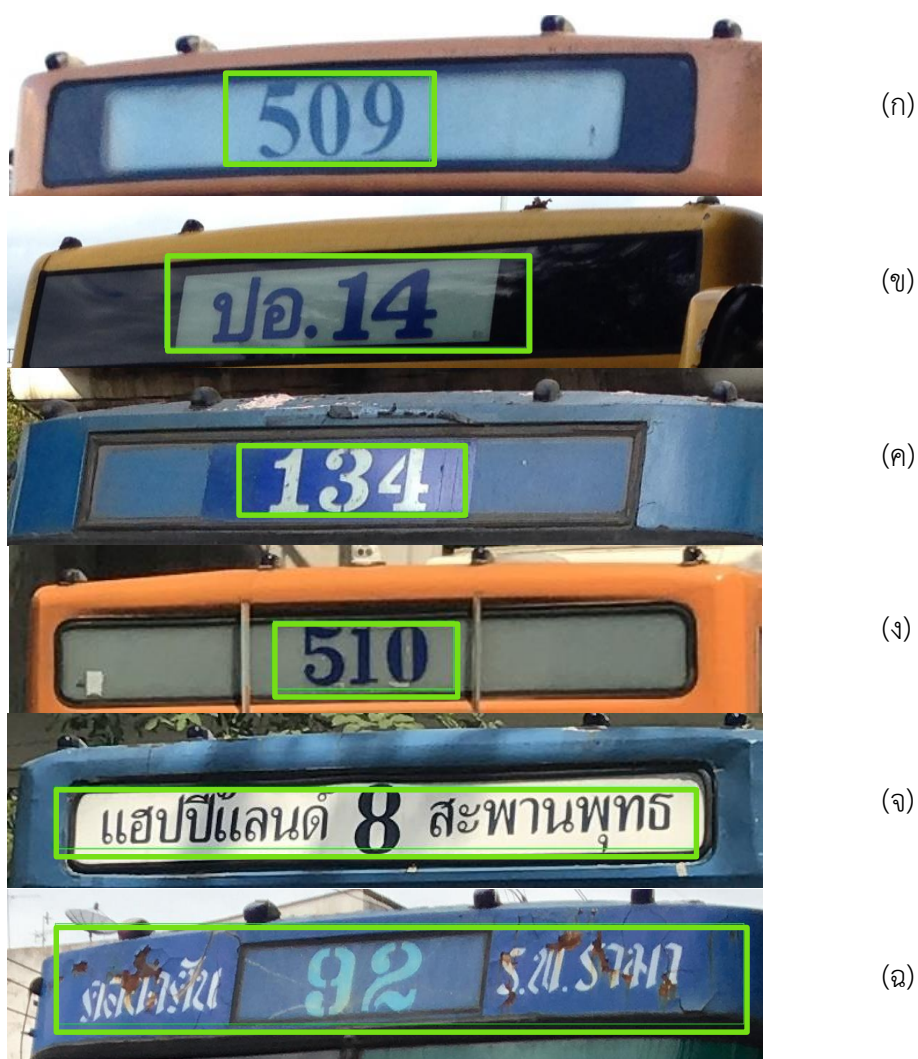


(ค)

รูปที่ 4.5 ตัวอย่างผลการทดลองของตัวตรวจจับบริเวณแผงด้านบนของรถประจำทางที่ผิดพลาด สีฟ้า คือบริเวณที่เป็นค่าความจริง, สีเขียว คือบริเวณที่ตรวจจับได้ (ก)(ข) ตัวตรวจจับ ตรวจจับได้ นอกเหนือบริเวณที่เป็นแผงด้านบนรถประจำทาง (ค) ตรวจจับไม่พบบริเวณแผงด้านบนรถประจำทาง

#### 4.3.2 การทดลองตัวตรวจหาบริเวณที่มีข้อความ

การทดลองที่สอง เป็นการทดลองตัวตรวจหาบริเวณที่มีข้อความ เมื่อได้บริเวณที่เป็นแผงด้านบนของรถประจำทาง จะนำผลลัพธ์ที่ได้มาเข้าสู่กระบวนการตรวจหาบริเวณที่มีข้อความ ที่ให้ผลลัพธ์เป็นพิกัดบริเวณที่เป็นข้อความ ซึ่งบริเวณที่เป็นข้อความนั้น หากบริเวณแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทางเพียงอย่างเดียว หรือบริเวณแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทางและข้อมูลสถานที่ที่ผ่าน หรือบริเวณแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ จะถือว่าเป็น 1 บริเวณข้อความ ดังแสดงรูปที่ 4.6



รูปที่ 4.6 ผลการทดลองของตัวตรวจหาบริเวณที่มีข้อความ (ก)(ค)(ง) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหาบริเวณที่มีข้อความ ซึ่งแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทางเพียงอย่างเดียว (ข) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหาบริเวณที่มีข้อความ ซึ่งแผงด้านบนรถประจำทางแสดงเลข

สายรถประจำทางกับตัวอักษรที่มีความหมายพิเศษ (จ)(ฉ) ผลลัพธ์ 1 บริเวณที่ได้จากการตรวจหา บริเวณที่มีข้อความ ซึ่งแผงด้านบนรถประจำทางแสดงเลขสายรถประจำทางและข้อมูลสถานที่ที่ผ่าน

จากภาพบริเวณแผงด้านบนของรถประจำทาง จำนวน 112 บริเวณ ผลที่ได้จากตัวตรวจหา บริเวณที่มีข้อความทำการตอบออกมาทั้งหมด 132 บริเวณ ถูกต้องจำนวน 107 บริเวณ โดยความ ผิดพลาด ถูกแบ่งออกเป็น 2 กลุ่ม กลุ่มแรกคือ ตัวตรวจหาบริเวณที่มีข้อความตรวจจับออกมาผิด จำนวน 25 บริเวณ และกลุ่มที่สอง ตัวตรวจหาบริเวณที่มีข้อความไม่สามารถตรวจจับข้อความได้ จำนวน 5 บริเวณ โดยผลการทดสอบแสดงดังตารางที่ 4.5

ตารางที่ 4.5 ผลการทดลองตัวตรวจหาบริเวณที่มีข้อความ

บริเวณแผงด้านบน ของรถประจำทาง นำเข้า (บริเวณ)	ผลลัพธ์จากตัวตรวจหาบริเวณที่มีข้อความ			
	ตัวตรวจหาบริเวณที่มี ข้อความตอบบริเวณ ที่มีข้อความ รวม ทั้งหมด (บริเวณ)	ถูกต้อง (บริเวณ) (TP)	ตรวจจับผิด (บริเวณ) (FP)	ตรวจจับไม่เจอ (บริเวณ) (FN)
112	132	107	25	5

รูปที่ 4.6 แสดงถึงภาพจากผลจากการทดลองในขั้นตอนของการตรวจหาบริเวณที่มีข้อความ ที่ตรวจหาได้ถูกต้อง และนำผลลัพธ์ไปเข้าสู่กระบวนการรู้จำข้อความต่อไป

และผลลัพธ์ที่ได้จากตัวตรวจหาบริเวณที่มีข้อความ ถูกประเมินประสิทธิภาพออกมาในค่า ความเที่ยง (Precision) ค่ารีคอล (Recall) และค่าการวัดเอฟ (F-Measure) ดังตารางที่ 4.6

ตารางที่ 4.6 การประเมินประสิทธิภาพของตัวตรวจหาบริเวณที่มีข้อความ

ตัวตรวจหาบริเวณที่มีข้อความ		
ค่าความเที่ยง	ค่ารีคอล	ค่าการวัดเอฟ
0.810	0.955	0.877

จากการทดลองพบว่า มีความผิดพลาดจากการตรวจหาบริเวณที่มีข้อความในบางภาพซึ่ง สามารถวิเคราะห์ได้เป็น 2 กรณี กรณีแรกคือ ภาพบริเวณแผงด้านบนของรถประจำทางนำเข้านั้น มี ข้อความอื่นๆ นอกจากข้อความที่เป็นข้อมูลสายรถประจำทาง และเลขสายรถประจำทางปะปนอยู่

หรือ สภาพแวดล้อม สิ่งของที่มีลักษณะที่คล้ายกับตัวอักษรหรือข้อความ ทำให้ตัวตรวจหาบริเวณที่มีข้อความทำนายว่าบริเวณนั้นมีข้อความหรือตัวอักษรปรากฏอยู่ รูปที่ 4.7 แสดงถึงตัวอย่างความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีแรก



รูปที่ 4.7 ตัวอย่างความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีที่ฝั่งด้านบนของรถประจำทางมีข้อความอื่นๆ นอกจากข้อความที่เป็นข้อมูลสายรถประจำทาง หรือ สภาพแวดล้อม สิ่งของที่มีลักษณะที่คล้ายกับตัวอักษรหรือข้อความ (ก)(จ) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความอื่นนอกจากเลขสายรถประจำทาง (ข)(ง) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความจากป้ายโฆษณา (ค) ตัวตรวจหาบริเวณที่มีข้อความตรวจเจอข้อความที่ควรเป็น 1 บริเวณ แต่ตรวจจับได้ 2 บริเวณ และอีก 1 บริเวณที่ตรวจจับผิด

ความผิดพลาดของตัวตรวจหาบริเวณที่มีข้อความในกรณีที่สอง คือ ไม่สามารถตรวจจับตัวอักษรได้ จากผลการทดลองของการตรวจหาบริเวณที่มีข้อความพบว่า มี 5 ภาพที่ไม่สามารถตรวจพบข้อความหรือตัวอักษร โดยจาก 4 ใน 5 บริเวณ เป็นภาพที่ตัวเลขรถประจำทางเป็นเลขตัวเดียว

จากการวิเคราะห์ ผู้วิจัยพบว่าเนื่องจากตัวตรวจหาบริเวณที่มีข้อความส่วนใหญ่ รวมทั้งตัวตรวจหาบริเวณที่มีข้อความที่ใช้ในวิทยานิพนธ์ฉบับนี้ ใช้สมมุติฐานในการตรวจหาว่า ข้อความจะประกอบด้วยตัวอักษรหลายตัวเรียงต่อกัน จึงทำให้ไม่สามารถตรวจจับตัวอักษรตัวเดียวได้ เพราะขนาดความยาวของข้อความที่มีความยาวที่สั้น ซึ่งถูกมองว่าเป็นสิ่งรบกวน โดยปกติหากเป็นบริเวณที่มีข้อความจะมีความยาวที่ยาวพอควร จึงสามารถตรวจจับได้ ตัวอย่างของความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีที่สอง แสดงดัง

รูปที่ 4.8



รูปที่ 4.8 ตัวอย่างความผิดพลาดของการตรวจหาบริเวณที่มีข้อความในกรณีที่ไม่สามารถตรวจจับตัวอักษรได้ (ก)(ข)(ค)(ง) คือภาพที่มีลักษณะเป็นตัวเลขเพียงตัวเดียวซึ่งตัวตรวจหาบริเวณที่มีข้อความไม่สามารถตรวจจับได้เนื่องจากสมมุติฐานของตัวตรวจจับข้อความ คือข้อความจะประกอบด้วยตัวอักษรหลายตัวเรียงต่อกัน (ง) ตัวตรวจจับข้อความตรวจจับผิดพลาด

### 4.3.3 การทดลองตัวรู้จำข้อความ

การทดลองที่สามเป็นการทดลองตัวรู้จำข้อความ โดยการนำผลลัพธ์ที่ได้จากการตรวจหาบริเวณที่มีข้อความมาเข้าสู่กระบวนการรู้จำข้อความจากกูเกิล คลาวด์วิชั่น ในลำดับแรก เนื่องจากการรู้จำข้อความจาก กูเกิล คลาวด์วิชั่นสามารถรู้จำภาษาไทยได้ ซึ่งเหมาะแก่การนำมารู้จำข้อความบนบริเวณแผงด้านบนรถประจำทางที่มีภาษาไทย หากคำตอบของกูเกิล คลาวด์วิชั่น ที่ตอบออกมาได้ผลลัพธ์เป็นไม่มีคำตอบ จะส่งภาพนำเข้าไปที่ไม่มีคำตอบนั้นไปเข้ากระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาวที่ใช้หลักการของการเลื่อนวินโดว์และนำไปเข้าสู่ตัวสกัดคุณลักษณะและจำแนกออกมาเป็นตัวเลขและตัวอักษรภาษาอังกฤษ ซึ่งสามารถรู้จำตัวเลขและตัวอักษรได้ดีกว่าตัวรู้จำข้อความของกูเกิลคลาวด์วิชั่น แต่ไม่สามารถรู้จำตัวอักษรภาษาไทยได้

การทดลองตัวรู้จำข้อความ ผู้วิจัยทำการจำแนกประเภทของคำตอบออกเป็น 6 กลุ่ม คือ

1. ถูกต้อง คือ ตัวรู้จำให้คำตอบตรงกับค่าความจริง หรือหมายความว่า  $(A \cap G = A)$  เช่น ค่าความจริง คือ “78” ตัวรู้จำข้อความตอบออกมาเป็น “78”
2. ครอบคลุม คือ ตัวรู้จำให้คำตอบซึ่งครอบคลุมค่าความจริง แต่มีตัวหนังสืออื่นปะปน หรือหมายความว่า  $(A \supset G)$  เช่น ค่าความจริง คือ “8” ตัวรู้จำข้อความตอบออกมาเป็น “ท8”
3. โกล้เคียง คือ ตัวรู้จำให้คำตอบใกล้เคียงกับค่าความจริง หรือหมายความว่า  $(A \cap G \neq \emptyset)$  เช่น ค่าความจริงคือ “ปอ14” ตัวรู้จำข้อความตอบเป็น “บอ14”
4. บางส่วน คือ ตัวรู้จำให้คำตอบได้บางส่วนของค่าความจริง หรือหมายความว่า  $(A \subset G)$  เช่น ค่าความจริงคือ “543” ตัวรู้จำข้อความตอบเป็น “43” หรือ “54” เป็นต้น
5. ไม่ถูกต้อง คือ ตัวรู้จำข้อความตอบออกมาได้ ไม่ตรงหรือไม่ใกล้เคียงกับค่าความจริง หรือหมายความว่า  $(A \cap G = \emptyset)$  เช่น ค่าความจริงคือ “543” ตัวรู้จำข้อความตอบเป็น “111”
6. ไม่มีคำตอบ คือ ตัวรู้จำข้อความไม่สามารถให้คำตอบใดๆ ออกมา หรือหมายความว่า  $N/A$

โดยที่  $A$  คือตัวเลขที่ทำนายได้จากการรู้จำ และ  $G$  คือตัวเลขค่าความจริง

จากบริเวณภาพที่ได้จากระบบการตรวจหาบริเวณที่มีข้อความ จำนวน 107 บริเวณ นำมาเข้ากระบวนการรู้จำข้อความจากกูเกิลคลาวด์วิชั่น ผลลัพธ์ที่ได้หลังจากทำการจำแนกประเภทคำตอบแสดงดังตารางที่ 4.7

ตารางที่ 4.7 ผลลัพธ์คำตอบที่ได้จากตัวรู้จำภูเกิลคลาวด์วิชั่น โดยจำแนกคำตอบเป็น 6 กลุ่ม

ผลลัพธ์คำตอบที่ได้จากตัวรู้จำ ภูเกิลคลาวด์วิชั่น						
จำนวน บริเวณภาพ	ถูกต้อง	ครอบคลุม	ใกล้เคียง	บางส่วน	ไม่ถูกต้อง	ไม่มีคำตอบ
107	53	9	8	2	1	34

จากตารางผลลัพธ์คำตอบที่ได้จากตัวรู้จำภูเกิลคลาวด์วิชั่น คำตอบที่ไม่มีคำตอบมีทั้งหมด 34 บริเวณภาพ จึงนำทั้ง 34 บริเวณภาพ เข้ากระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาวอีกครั้ง เพื่อเพิ่มความถูกต้องให้กับผลลัพธ์คำตอบ โดยผลลัพธ์ทั้ง 34 บริเวณภาพ เมื่อนำเข้ากระบวนการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาวจะได้ผลลัพธ์ออกมาดังตารางที่ 4.8

ตารางที่ 4.8 ผลลัพธ์คำตอบ จากผลลัพธ์ไม่มีคำตอบของภูเกิลคลาวด์วิชั่น เมื่อเข้าสู่กระบวนการการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว

ผลลัพธ์คำตอบที่ได้จากผลลัพธ์ไม่มีคำตอบของภูเกิลคลาวด์วิชั่น นำเข้าสู่กระบวนการการรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว						
จำนวน บริเวณภาพ	ถูกต้อง	ครอบคลุม	ใกล้เคียง	บางส่วน	ไม่ถูกต้อง	ไม่มีคำตอบ
34	7	3	13	0	11	0

เมื่อทำการรวมผลลัพธ์ทั้งหมดของขั้นตอนการรู้จำข้อความ ทั้งตัวรู้จำข้อความภูเกิลคลาวด์วิชั่น และตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว ผลลัพธ์ที่ได้สุดท้ายแสดงดังตารางที่ 4.9

ตารางที่ 4.9 ผลลัพธ์คำตอบโดยรวมทั้งหมดของขั้นตอนการรู้จำข้อความ

ผลลัพธ์คำตอบโดยรวมทั้งหมดของขั้นตอนการรู้จำข้อความ						
จำนวน บริเวณภาพ	ถูกต้อง	ครอบคลุม	ใกล้เคียง	บางส่วน	ไม่ถูกต้อง	ไม่มีคำตอบ
107	60	12	21	2	12	0

ผลลัพธ์คำตอบโดยรวมทั้งหมดของขั้นตอนการรู้จำข้อความ สามารถรู้จำข้อความได้ถูกต้อง จำนวน 60 บริเวณภาพ, ครอบคลุม 12 บริเวณภาพ, โกล้เคียง 21 บริเวณภาพ, บางส่วน 2 บริเวณภาพ, ไม่ถูกต้อง 12 บริเวณภาพ ซึ่งนำผลลัพธ์คำตอบโดยรวมทั้งหมดนี้ เข้าสู่กระบวนการการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ เพื่อทำการแก้ไขคำตอบที่ยังผิดหรือคำตอบที่ยังถูกต้องแบบมีเงื่อนไข ให้สามารถกลับมาถูกต้องได้

#### 4.3.4 การทดลองหลังจากการทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ

การทดลองที่สี่เป็นการทดลองหลังจากการทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ โดยการนำผลลัพธ์ที่ได้จากขั้นตอนการรู้จำข้อความมาทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ โดยขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำประกอบไปด้วย 2 ขั้นตอน คือ ขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง ที่ทำการตัดตัวอักษรทุกตัวด้วยเว้น 0-9, ปอ, ป, อ, ก, A, a, S, s และขั้นตอนการใช้ข้อมูลจีพีเอสของป้ายรถประจำทาง ในการช่วยแก้ไขคำตอบที่ผิด ให้ถูกต้องมากขึ้น

ผลลัพธ์ที่ได้หลังจากการทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำในส่วนขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง จะทำการแยกคำตอบเป็น 6 กลุ่ม เหมือนกับขั้นตอนการรู้จำข้อความ โดยแสดงดังตารางที่ 4.10

ตารางที่ 4.10 ผลลัพธ์คำตอบจากขั้นตอนการประมวลผลหลังในส่วนขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง

ผลลัพธ์คำตอบจากขั้นตอนการประมวลผลหลังในส่วนขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง						
จำนวนบริเวณภาพ	ถูกต้อง	ครอบคลุม	ใกล้เคียง	บางส่วน	ไม่ถูกต้อง	ไม่มีคำตอบ
107	66	8	14	7	10	2

หลังจากการตัดตัวอักษรที่ไม่เกี่ยวข้อง ทำให้ผลลัพธ์มีความถูกต้องเพิ่มขึ้น จากนั้นนำผลลัพธ์หลังจากการตัดตัวอักษรที่ไม่เกี่ยวข้องเข้าสู่ขั้นตอนที่สอง คือการใช้ข้อมูลจีพีเอสของป้ายรถประจำทางมาช่วยในการแก้ไขคำตอบ

ตารางที่ 4.11 คือตารางคอนฟิวชันเมตริกซ์ เปรียบเทียบระหว่างผลลัพธ์คำตอบจากขั้นตอนในการตัดตัวอักษรที่ไม่เกี่ยวข้อง และผลลัพธ์คำตอบหลังจากใช้ข้อมูลจีพีเอสของป้ายรถประจำทางในการช่วยแก้ไขข้อมูล จากตารางเห็นว่าการใช้ข้อมูลจีพีเอสของป้ายรถประจำทางในส่วนข้อมูลเลขสายรถประจำทางที่ผ่านป้าย สามารถช่วยให้ผลลัพธ์คำตอบในกลุ่มของคำตอบ ครอบคลุม, โกล้เคียง,



บางส่วน กลับเป็นคำตอบที่ถูกต้อง โดยผลลัพธ์ตอบเลขสายรถประจำทางถูกต้องทั้งหมด 70 บริเวณภาพ จากภาพนำเข้าไปในกระบวนการประมวลผลหลังทั้งหมด ส่วนของการตัดตัวอักษรที่ไม่เกี่ยวข้อง จำนวน 107 บริเวณ

**ตารางที่ 4.11** ตารางคอนฟิวชันเมทริกซ์เปรียบเทียบระหว่างผลลัพธ์คำตอบจากขั้นตอนในการตัดตัวอักษรที่ไม่เกี่ยวข้อง และผลลัพธ์คำตอบหลังจากใช้ข้อมูลจีพีเอสของป้ายรถประจำทาง

		ผลลัพธ์คำตอบจากขั้นตอนการประมวลผลหลังในส่วนขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง โดยใช้ข้อมูลจีพีเอสของป้ายรถประจำทาง						
		ถูกต้อง	ครอบคลุม	ใกล้เคียง	บางส่วน	ไม่ถูกต้อง	ไม่มีคำตอบ	รวม
คำตอบจากขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง	ถูกต้อง	63	0	3	0	0	0	66
	ครอบคลุม	4	4	0	0	0	0	8
	ใกล้เคียง	0	0	11	3	0	0	14
	บางส่วน	3	0	2	2	0	0	7
	ไม่ถูกต้อง	0	0	0	0	10	0	10
	ไม่มีคำตอบ	0	0	0	0	0	2	2
	รวม	70	4	16	5	10	2	107

หลังจากได้ผลลัพธ์คำตอบจากการใช้ข้อมูลจีพีเอสของป้ายรถประจำทางช่วยในการแก้ไขคำตอบ ส่วนของคำตอบที่ยังคงอยู่ในกลุ่มของคำตอบที่ ครอบคลุม, ใกล้เคียง, บางส่วน และ ไม่มีคำตอบที่มีจำนวนรวม 37 บริเวณภาพ มีสาเหตุเกิดจาก การรู้จำข้อความ รู้จำผิด ทำให้ไม่สามารถเปรียบเทียบคำตอบสายรถประจำทางที่ใกล้เคียงที่มาจากข้อมูลจีพีเอสของป้ายรถประจำทางได้ จำนวน 26 บริเวณภาพ, ข้อมูลของป้ายรถประจำทาง มีข้อมูลของเลขสายรถประจำทางที่ผ่าน ไม่ครบถ้วน จำนวน 8 บริเวณภาพ ภาพถ่ายนำเข้าไปไม่มีข้อมูลจีพีเอสและรู้จำข้อความผิด จำนวน 1 บริเวณภาพ และหลังจากขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ในส่วนการตัดตัวอักษรที่ไม่เกี่ยวข้อง คำตอบถูกต้องตัดตัวอักษรที่ไม่เกี่ยวข้องทั้งหมด จึงไม่มีข้อมูลที่นำมาเปรียบเทียบได้ จำนวน 2 บริเวณภาพ

#### 4.3.5 รวมผลการทดลองทั้งโปรแกรม

เมื่อนำการทดลองทั้งหมดมารวมกัน ทั้งขั้นตอนการหาบริเวณแฉงด้านบนรถประจำทาง ขั้นตอนการหาบริเวณที่มีตัวหนังสือ ขั้นตอนการรู้จำข้อความ และขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ จากภาพนำเข้าจำนวน 108 ภาพ สามารถหาบริเวณที่เป็นแฉงด้านบนของรถประจำทางได้จำนวน 112 บริเวณ ส่งต่อเข้าไปยังขั้นตอนหาบริเวณที่มีข้อความ ขั้นตอนรู้จำข้อความ และขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ผลลัพธ์สุดท้ายที่ได้คือตอบเป็นเลขสายรถประจำทางถูกต้องจำนวน 70 บริเวณภาพ, ครอบคลุม 4 บริเวณภาพ, ใกล้เคียง 16 บริเวณภาพ, บางส่วน 5 บริเวณภาพ, ไม่ถูกต้อง 10 บริเวณภาพ และไม่มีคำตอบจำนวน 7 บริเวณภาพ แสดงดังตารางที่ 4.12 หากนำผลลัพธ์สุดท้ายที่ตอบถูกต้องจำนวน 70 บริเวณภาพ มาคำนวณร่วมกับค่าความจริงของบริเวณที่เป็นแฉงด้านบนรถประจำทางจำนวน 113 บริเวณ จะได้เปอร์เซ็นต์ความแม่นยำทั้งหมด 62 เปอร์เซ็นต์

ตารางที่ 4.12 ตารางแสดงผลลัพธ์ของการทดลองทั้งโปรแกรม

ผลการทดลองทั้งโปรแกรม								
จำนวนภาพ แฉงด้านบน รถประจำ ทางทั้งหมด (บริเวณ)	จำนวนภาพ แฉงด้านบนรถ ประจำทางที่ ตรวจจับ ถูกต้อง (บริเวณ)	ถูกต้อง	ครอบ คลุม	ใกล้ เคียง	บาง ส่วน	ไม่ ถูกต้อง	ไม่มี คำตอบ	รวม
113	112	70	4	16	5	10	7	112

#### 4.4 สรุปผลการทดลอง

การทดลองของระบบอ่านเลขสายรถประจำทางจากภาพประกอบไปด้วยการทดลองหลัก 4 การทดลอง คือ การทดลองขั้นตอนการหาบริเวณแฉงด้านบนของรถประจำทาง การทดลองขั้นตอนการหาบริเวณที่มีข้อความ การทดลองขั้นตอนการรู้จำข้อความ และการทดลองขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ

#### 4.4.1 การทดลองขั้นตอนการหาบริเวณแผงด้านบนของรถประจำทาง

จากภาพนำเข้าจำนวน 108 ภาพ นำเข้าสู่ขั้นตอนการหาแผงด้านบนของรถประจำทางโดยผลลัพธ์คำตอบที่ได้ ตอบบริเวณแผงด้านบนรถประจำทางถูกจำนวน 112 บริเวณ โดยจากผลการทำนายว่าเป็นบริเวณแผงหน้ารถประจำทางนั้น ถูกต้องกับค่าจริงร้อยละ หรือค่าความเที่ยง (Precision) 0.982, จากจำนวนบริเวณแผงหน้ารถประจำทางจริงทั้งหมดนั้น ระบบทำนายถูกต้องร้อยละ หรือค่ารีคอล (Recall) 0.991 และค่าการวัดเอฟ (F-Measure) ซึ่งเป็นการวัดประสิทธิภาพโดยรวมของค่าความเที่ยงและค่ารีคอล มีค่า 0.986

#### 4.4.2 การทดลองขั้นตอนการหาบริเวณที่มีข้อความ

นำผลลัพธ์คำตอบบริเวณแผงด้านบนรถประจำทางเข้าสู่ขั้นตอนการหาบริเวณที่มีข้อความผลลัพธ์ที่ตอบถูกจากขั้นตอนการหาบริเวณที่มีข้อความมีจำนวน 107 บริเวณ โดยจากผลการทำนายว่าเป็นบริเวณที่มีข้อความนั้น ถูกต้องกับค่าจริงร้อยละ หรือค่าความเที่ยง (Precision) 0.810, จากจำนวนบริเวณที่มีข้อความจริงทั้งหมดนั้น ระบบทำนายถูกต้องร้อยละ หรือค่ารีคอล (Recall) 0.955 และค่าการวัดเอฟ (F-Measure) ซึ่งเป็นการวัดประสิทธิภาพโดยรวมของค่าความเที่ยงและค่ารีคอล มีค่า 0.877

#### 4.4.3 การทดลองขั้นตอนการรู้จำข้อความ

นำผลลัพธ์จากขั้นตอนการหาบริเวณที่มีข้อความเข้าสู่ขั้นตอนการรู้จำข้อความจำนวน 107 บริเวณ โดยเข้าตัวรู้จำข้อความของ กูเกิล คลาวด์วิชันเป็นลำดับแรก และหากผลลัพธ์ที่ได้จากกูเกิล คลาวด์วิชัน ตอบเป็นไม่มีคำตอบ ภาพนำเข้านั้นจะถูกส่งไปยังตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้นระยะยาว ซึ่งสุดท้ายผลลัพธ์ที่ได้ถูกจำแนกข้อมูลออกเป็น 6 กลุ่ม โดยได้ผลลัพธ์คำตอบที่ถูกต้องจำนวน 60 บริเวณภาพ, ครอบคลุม 12 บริเวณภาพ, ใกล้เคียง 21 บริเวณภาพ, บางส่วน 2 บริเวณภาพ, ไม่ถูกต้อง 12 บริเวณภาพ และไม่มีคำตอบ 0 บริเวณภาพ

#### 4.4.4 การทดลองขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ

นำคำตอบจากขั้นตอนการรู้จำข้อความเข้าสู่ขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ โดยการประมวลผลหลังเพื่อปรับปรุงความแม่นยำประกอบไปด้วย 2 ขั้นตอน คือ ขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง และขั้นตอนการใช้ข้อมูลจีพีเอสของป้ายรถประจำทาง ซึ่งการประมวลผลหลังเพื่อปรับปรุงความแม่นยำสามารถช่วยให้คำตอบที่ตอบผิด หรือคำตอบที่อยู่ในกลุ่มของครอบคลุม

ใกล้เคียง บางส่วน ตอบถูกต้องมากขึ้น ผลลัพธ์สุดท้ายที่ได้หลังจากการทำการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ได้ผลลัพธ์ที่ตอบเลขสายรถประจำทางถูกต้องจำนวน 70 บริเวณภาพ, ครอบคลุม 4 บริเวณภาพ, ใกล้เคียง 16 บริเวณภาพ, บางส่วน 5 บริเวณภาพ, ไม่ถูกต้อง 10 บริเวณภาพ และไม่มีคำตอบจำนวน 2 บริเวณภาพ

#### 4.4.5 รวมผลการทดลองทั้งโปรแกรม

โปรแกรมนี้จะสมบูรณ์ได้เมื่อนำการทดลองทั้งหมดมารวมกัน โดยวัดผลจากภาพนำเข้าที่ใช้ในการทดลองจำนวน 108 ภาพ มีค่าความจริงที่เป็นบริเวณแฉกด้านบนรถประจำทางจำนวน 113 บริเวณ ซึ่งตรวจจับแฉกด้านบนรถประจำทางถูกต้องจำนวน 112 บริเวณ เมื่อนำภาพแฉกด้านบนรถประจำทางเข้าสู่กระบวนการหาเลขสายรถประจำทาง ผลลัพธ์สุดท้ายตอบเลขสายรถประจำทางถูกต้องจำนวน 70 บริเวณภาพ, ครอบคลุม 4 บริเวณภาพ, ใกล้เคียง 16 บริเวณภาพ, บางส่วน 5 บริเวณภาพ, ไม่ถูกต้อง 10 บริเวณภาพ และไม่มีคำตอบจำนวน 7 บริเวณภาพ หากนำผลลัพธ์สุดท้ายที่ตอบถูกต้องจำนวน 70 ภาพ มาคำนวณร่วมกับค่าความจริงของบริเวณที่เป็นแฉกด้านบนรถประจำทางจำนวน 113 บริเวณ จะได้เปอร์เซ็นต์ความแม่นยำทั้งหมด 62 เปอร์เซ็นต์

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

บทนี้จะสรุปผลการวิจัยทั้งหมด วิธีการและผลการทดลองที่ได้นำเสนอ รวมถึงข้อเสนอแนะที่ควรปรับปรุงแก้ไข โดยมีรายละเอียดดังนี้

#### 5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอระบบอ่านเลขสายรถประจำทางจากภาพที่ถ่ายจากกล้องสมาร์ตโฟน โดยมีแรงบันดาลใจจากความต้องการที่จะช่วยเหลือผู้พิการสายตาเลือนรางให้สามารถโดยสารรถประจำทางได้อย่างถูกต้อง เนื่องจากผู้พิการสายตาเลือนรางมักมีความยากลำบากในการเดินทางไปยังสถานที่ต่างๆ เนื่องจากมองเห็นไม่ชัด หรือมองเห็นแคบ จึงมักจำเป็นต้องมีผู้ช่วยในการเดินทางไปยังสถานที่ต่างๆ แต่หากผู้พิการสายตาเลือนรางต้องการเดินทางด้วยตนเอง หรือหากอยู่ลำพังและต้องการใช้บริการรถประจำทาง ระบบสามารถช่วยผู้พิการสายตาเลือนรางให้ทราบถึงสายรถประจำทางได้

ระบบอ่านเลขสายรถประจำทางนี้ใช้วิธีการเรียนรู้ของเครื่อง เทคนิคการประมวลผลภาพ และใช้ข้อมูลที่เกี่ยวข้องประกอบ ข้อมูลภาพที่ใช้ทดสอบในงานวิจัยนี้ เป็นภาพถ่ายของรถประจำทางของ ขสมก. หรือรถร่วมบริการที่วิ่งบริการอยู่ในกรุงเทพมหานครและปริมณฑล โดยภาพถูกถ่ายบริเวณป้ายรถประจำทางในเวลากลางวัน ด้วยจำนวนผู้ถ่ายภาพทั้งหมด 5 คน เพื่อความหลากหลายของมุมมองรถประจำทาง จากนั้นนำเข้าสู่ขั้นตอนการวิเคราะห์เพื่อหาเลขสายรถประจำทาง โดยมีทั้งหมด 4 ขั้นตอนหลัก คือ ขั้นตอนการหาแผงด้านบนของรถประจำทาง ขั้นตอนการหาบริเวณที่มีข้อความ ขั้นตอนการรู้จำข้อความ และการประมวลผลภายหลังเพื่อปรับปรุงความแม่นยำ

ขั้นตอนการหาแผงด้านบนของรถประจำทางใช้วิธีการการเรียนรู้ของเครื่องในส่วนของ การเรียนรู้เชิงลึกที่สามารถรู้คุณลักษณะสำคัญจากชุดข้อมูลที่สอนโดยวิธีการคอนโวลูชันนอลนิรอลเน็ตเวิร์คแบบฟาสเตอร์อาร์ชีเอ็นเอ็น โดยต้องเตรียมชุดข้อมูลสอนที่เป็นพิกัดบริเวณแผงด้านบนของรถประจำทางเพื่อทำการสอน โดยผลลัพธ์บริเวณแผงด้านบนรถประจำทางที่ได้ ตอบถูกจำนวน 112 บริเวณ และมีค่าความเที่ยง (Precision) 0.982, ค่ารีคอล (Recall) 0.991 และค่าการวัดเอฟ (F-Measure) 0.986

ขั้นตอนการหาบริเวณที่มีข้อความ ได้เลือกใช้ตัวตรวจหาบริเวณที่มีข้อความที่ใช้กระบวนการคอนโวลูชันนอลนิรอลเน็ตเวิร์คแบบฟาสเตอร์อาร์ชีเอ็นเอ็น ที่ถูกสอนจากชุดข้อมูลภาพตัวอักษรโดยมีทั้งภาษาไทย ภาษาอังกฤษ และตัวเลข จากฐานข้อมูลของอิมเมจเน็ต (ImageNet) และไอซีดีเออาร์

(ICDAR) โดยผลลัพธ์ที่ได้ ตอบออกมาเป็นพิกัดของบริเวณที่มีข้อความ ซึ่งผลการทดสอบการตรวจหา บริเวณที่มีข้อความตอบถูกต้องจำนวน 107 บริเวณ และมีค่าความเที่ยง (Precision) 0.810, ค่ารีคอล (Recall) 0.955 และค่าการวัดเอฟ (F-Measure) 0.877

ขั้นตอนการรู้จำข้อความ ได้เลือกใช้ตัวรู้จำข้อความจากกูเกิลคลาวด์วิชั่น ในการรู้จำผลลัพธ์ที่ได้จากขั้นตอนการหาบริเวณที่มีข้อความ หากผลลัพธ์คำตอบจากตัวรู้จำข้อความกูเกิลคลาวด์วิชั่น ตอบเป็นไม่มีคำตอบ ภาพที่ไม่มีคำตอบจะถูกนำไปยังตัวรู้จำข้อความด้วยวิธีหน่วยความจำระยะสั้น ระยะยาว เพื่อทำการรู้จำข้อความอีกครั้ง ผลลัพธ์สุดท้าย ถูกจำแนกข้อมูลออกเป็น 6 กลุ่ม โดยได้ผลลัพธ์คำตอบที่เป็นเลขสายรถประจำทางถูกต้องจำนวน 60 บริเวณภาพ ครอบคลุม 12 บริเวณภาพ ใกล้เคียง 21 บริเวณภาพ บางส่วน 2 บริเวณภาพ ไม่ถูกต้อง 12 บริเวณภาพ ไม่มีคำตอบ 0 บริเวณภาพ

ขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ จะนำคำตอบที่ได้จากขั้นตอนการรู้จำข้อความ เข้าสู่การประมวลผลหลังเพื่อแก้ไขคำตอบให้ถูกต้องมากขึ้น โดยขั้นตอนการประมวลผลหลัง เพื่อปรับปรุงความแม่นยำประกอบด้วย 2 ขั้นตอนคือ ขั้นตอนการตัดตัวอักษรที่ไม่เกี่ยวข้อง และ ขั้นตอนการใช้ข้อมูลจีพีเอสของป้ายรถประจำทางช่วยแก้ไขคำตอบ ผลลัพธ์สุดท้ายที่ได้หลังจากการทำ การประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ได้ผลลัพธ์ที่เป็นเลขสายรถประจำทางถูกต้องจำนวน 70 บริเวณภาพ ครอบคลุม 4 บริเวณภาพ ใกล้เคียง 16 บริเวณภาพ บางส่วน 5 บริเวณภาพ ไม่ถูกต้อง 10 บริเวณภาพ และ ไม่มีคำตอบจำนวน 2 บริเวณภาพ

จากขั้นตอนการทดลองทั้งหมด 4 ขั้นตอนพบว่า ขั้นตอนการตรวจจับบริเวณแผงด้านบนรถ ประจำทางยังมีข้อจำกัดในส่วนของรถประจำทางที่บริเวณเลขสายรถประจำทาง ไม่ได้อยู่บริเวณแผง ด้านบนของรถประจำทาง เนื่องจากการสอนเพื่อสร้างตัวตรวจจับแผงด้านบนของรถประจำทาง มี ข้อมูลภาพรถประจำทางที่มีลักษณะนี้น้อย จึงทำให้ไม่สามารถตรวจจับได้ และในส่วนขั้นตอนการรู้จำ ข้อความซึ่งเป็นขั้นตอนสำคัญ ยังมีข้อจำกัดในด้านต่างๆ เช่น ลักษณะของตัวหนังสือ ตัวหนังสือขาด หาย หรือสิ่งรบกวนภายในภาพ ทำให้ตัวรู้จำข้อความ รู้จำผิดพลาด ส่งผลกระทบต่อขั้นตอนในลำดับ ถัดไป และสุดท้ายขั้นตอนการประมวลผลหลังเพื่อปรับปรุงความแม่นยำ ในส่วนของข้อมูลจีพีเอสของ ป้ายรถประจำทาง ยังมีข้อมูลไม่ครบถ้วน ส่งผลทำให้ไม่สามารถนำมาช่วยในการแก้ไขผลลัพธ์ให้ ถูกต้องได้

## 5.2 ข้อเสนอแนะ

วิทยานิพนธ์ฉบับนี้ สามารถนำไปปรับปรุงและพัฒนาต่อยอดในอนาคตได้ โดยมีข้อเสนอแนะ และแนวทางดังนี้

### 5.2.1 ตัวตรวจจับบริเวณแผงด้านบนของรถประจำทาง

ตัวตรวจจับบริเวณแผงด้านบนของรถประจำทางที่ผู้วิจัยนำเสนอใช้กระบวนการการเรียนรู้เชิงลึกวิธีคอลโลลูชันนอลนิวโรลเน็ตเวิร์ค แบบฟาสเตอร์อาร์ซีเอ็นเอ็น ซึ่งเป็นวิธีที่ให้ความแม่นยำที่สูง แต่ใช้เวลาในการประมวลผลมาก ดังนั้นหากนำระบบนี้ไปพัฒนาลงบนสมาร์ตโฟน อาจต้องใช้วิธีการการเรียนรู้เชิงลึกในแบบอื่นที่ประมวลผลเร็ว แต่อาจต้องแลกด้วยความแม่นยำที่ต่ำ และควรทำการเพิ่มจำนวนของชุดข้อมูลสอน ให้มีความครอบคลุมมากขึ้น

### 5.2.2 ตัวตรวจจับบริเวณที่มีข้อความ

ตัวตรวจจับบริเวณที่มีข้อความที่ใช้เป็นตัวตรวจจับบริเวณที่มีข้อความที่ให้คำตอบพิกัดบริเวณที่มีข้อความเป็นรูปร่างสี่เหลี่ยมตรง หากตัวหนังสือจากแผงด้านบนของรถประจำทางมีลักษณะที่เอียงอาจทำให้ไม่สามารถตรวจจับได้ หรือหากตรวจจับได้ จะส่งผลต่อการรู้จำข้อความในขั้นตอนถัดไป จึงจำเป็นต้องพัฒนาตัวตรวจจับบริเวณที่มีข้อความ ให้ตรวจจับข้อความออกมาในลักษณะเอียงตามข้อความได้

### 5.2.3 ตัวรู้จำข้อความ

เนื่องจากแผงด้านบนของรถประจำทาง ส่วนมากมีตัวอักษรหรือข้อความที่เป็นภาษาไทยประกอบ จึงจำเป็นต้องมีตัวรู้จำข้อความภาษาไทยที่มีประสิทธิภาพในการรู้จำตัวอักษรหรือข้อความบริเวณแผงด้านบนรถประจำทาง

### 5.2.4. การประมวลผลหลังเพื่อปรับปรุงความแม่นยำ

การประมวลผลหลังเพื่อปรับปรุงความแม่นยำในส่วนข้อมูลจีพีเอสของป้ายรถประจำทางที่ใช้ยังมีข้อมูลของป้ายรถประจำทางและสายรถประจำทางที่ผ่านป้ายไม่ครบถ้วน ซึ่งข้อมูลจีพีเอสของป้ายรถประจำทางที่ใช้ในวิทยานิพนธ์ได้มาจากแผนที่ลองดู จึงอาจจำเป็นต้องทำการหาแหล่งข้อมูลจีพีเอสของป้ายรถประจำทางอื่นที่มีข้อมูลที่มาก หรือครบถ้วน เช่น ข้อมูลป้ายรถประจำทางของ ขสมก.

## รายการอ้างอิง

- [1] "ICD-10 Version:2016." [Online]. Available: <http://apps.who.int/classifications/icd10/browse/2016/en#/H54>. [Accessed: 27-Feb-2018].
- [2] "ตัวอย่างการมองเห็นของผู้พิการสายตาสีอ่อนราง." [Online]. Available: <http://www.lowvisionthailand.org>, <https://www.smashingmagazine.com/2014/10/color-contrast-tips-and-tools-for-accessibility/>, <http://www.inclusivedesigntoolkit.com/idt-cn/UCvision/vision.html>. [Accessed: 21-Jun-2016].
- [3] "ตัวอย่างลักษณะของรถประจำทางในกรุงเทพมหานคร." [Online]. Available: <http://www.medanbisnisdaily.com>. [Accessed: 21-Jun-2017].
- [4] "แนวคิด Neural Network." [Online]. Available: <https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic>. [Accessed: 21-May-2018].
- [5] "Efficient mapping of the training of Convolutional Neural Networks to a CUDA-based cluster." [Online]. Available: <http://parse.ele.tue.nl/education/cluster2>. [Accessed: 11-May-2016].
- [6] "Deep Learning for Object Detection: A Comprehensive Review." [Online]. Available: <https://towardsdatascience.com/deep-learning-for-object-detection-a-comprehensive-review-73930816d8d9>. [Accessed: 21-Feb-2018].
- [7] J. R. R. Uijlings, K. E. A. v. d. Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154-171, 2013.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.
- [9] "Deep Learning for Object Detection: A Comprehensive Reviews." [Online]. Available: <https://www.kdnuggets.com/2017/10/deep-learning-object-detection-comprehensive-review.html>. [Accessed: 14-Feb-2018].



- [10] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440-1448.
- [11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems 28*: Curran Associates, Inc., 2015, pp. 91-99.
- [12] "ImageNet." [Online]. Available: [www.image-net.org](http://www.image-net.org). [Accessed: 16-Feb-2018].
- [13] "COCO Dataset." [Online]. Available: [cocodataset.org](http://cocodataset.org). [Accessed: 16-Feb-2018].
- [14] "ImageNet: VGGNet, ResNet, Inception, and Xception with Keras." [Online]. Available: <https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/>. [Accessed: 17-Mar-2018].
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.
- [16] "Quick complete Tensorflow tutorial to understand and run Alexnet, VGG, Inceptionv3, Resnet and squeezeNet networks." [Online]. Available: <http://cv-tricks.com/tensorflow-tutorial/understanding-alexnet-resnet-squeezenetand-running-on-tensorflow/>. [Accessed: 21-Nov-2017].
- [17] "Understanding LSTM Networks." [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. [Accessed: 01-Apr-2018].
- [18] "การทำงานของหน่วยความจำระยะสั้นระยะยาว." [Online]. Available: <https://medium.com/@sinart.t/long-short-term-memory-lstm-e6cb23b494c6>. [Accessed: 21-Jun-2018].
- [19] "Haversine Formula." [Online]. Available: [https://en.wikipedia.org/wiki/Haversine\\_formula](https://en.wikipedia.org/wiki/Haversine_formula). [Accessed: 27-Feb-2018].
- [20] "ATAN2 (ฟังก์ชัน ATAN2)." [Online]. Available: <https://support.office.com/th-th/article/atan2-%E0%B8%9F%E0%B8%B1%E0%B8%87%E0%B8%81%E0%B9%8C%E0%B8%8A%E0%B8%B1%E0%B8%99-atan2-c04592ab-b9e3-4908-b428-c96b3a565033>. [Accessed: 1-Jul-2018].

- [21] "TensorFlow." [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 04-Jan-2018].
- [22] "Find minimum edit distance between given two strings." [Online]. Available: <https://www.ideserve.co.in/learn/edit-distance-dynamic-programming>. [Accessed: 1-Jul-2018].
- [23] "difflib — Helpers for computing deltas." [Online]. Available: <https://docs.python.org/3/library/difflib.html#difflib.SequenceMatcher>. [Accessed: 14-May-2018].
- [24] "Intersection over Union (IoU) for object detection." [Online]. Available: <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>. [Accessed: 04-July-2016].
- [25] "The PASCAL Visual Object Classes Homepage." [Online]. Available: <http://host.robots.ox.ac.uk/pascal/VOC/>. [Accessed: 01-May-2018].
- [26] "Precision and recall." [Online]. Available: [https://en.wikipedia.org/wiki/Precision\\_and\\_recall#Precision](https://en.wikipedia.org/wiki/Precision_and_recall#Precision). [Accessed: 01-May-2018].
- [27] M. Z. H. Noor, I. Ismail, and M. F. Saaid, "Bus detection device for the blind using RFID application," in *5th International Colloquium on Signal Processing Its Applications, 2009. CSPA 2009, 2009*, pp. 247-249.
- [28] J. A. Kalbani, R. B. Suwailam, A. A. Yafai, D. A. Abri, and M. Awadalla, "Bus detection system for blind people using RFID," in *GCC Conference and Exhibition (GCCCE), 2015 IEEE 8th, 2015*, pp. 1-6.
- [29] C. Guida, D. Comanducci, and C. Colombo, "Automatic Bus Line Number Localization and Recognition on Mobile Phones—A Computer Vision Aid for the Visually Impaired," in *Image Analysis and Processing – ICIAP 2011: Springer Berlin Heidelberg, 2011*, pp. 323-332.
- [30] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [31] H. Pan, C. Yi, and Y. Tian, "A primary travelling assistant system of bus detection and recognition for visually impaired people," in *2013 IEEE*

- International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1-6.
- [32] C. M. Tsai and Z. M. Yeh, "Text detection in bus panel for visually impaired people ??seeing?? bus route number," in *2013 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2013, vol. 03, pp. 1234-1239.
- [33] D. Lee, H. Yoon, C. Park, J. Kim, and C. H. Park, "Automatic number recognition for bus route information aid for the visually-impaired," in *2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 2013, pp. 280-284.
- [34] C.-M. Tsai and Z.-M. Yeh, "Detection of Bus Routes Number in Bus Panel via Learning Approach," in *Intelligent Information and Database Systems: Springer International Publishing*, 2014, pp. 302-311.
- [35] W. Ma, H. Yang, and Y. Wang, "A Robust Bus Detection and Recognition Method Based on 3D Model and LSD Method for Public Security Road Crossing Application," in *Advances on Digital Television and Wireless Multimedia Communications: Springer Berlin Heidelberg*, 2012, pp. 73-81.
- [36] C.-C. Cheng and C.-M. Tsai, "Using Red-Otsu Thresholding to Detect the Bus Routes Number for Helping Blinds to Take Bus," in *Modern Advances in Applied Intelligence: Springer International Publishing*, 2014, pp. 321-330.
- [37] "LabelImg." [Online]. Available: <https://github.com/tzutalin/labelImg/>. [Accessed: 01-May-2018].
- [38] "Raccoon Detector Dataset." [Online]. Available: [https://github.com/datitran/raccoon\\_dataset](https://github.com/datitran/raccoon_dataset). [Accessed: 02-May-2018].
- [39] T. Kobchaisawat and T. H. Chalidabhongse, "Thai text localization in natural scene images using Convolutional Neural Network," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, pp. 1-7.
- [40] T. Kobchaisawat and T. H. Chalidabhongse, "A method for multi-oriented Thai text localization in natural scene images using Convolutional Neural Network," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2015, pp. 220-225.

- [41] "ICDAR 2013 Robust Reading Competition." [Online]. Available: <http://dagdata.cvc.uab.es/icdar2013competition/>. [Accessed: 04-May-2018].
- [42] "Cloud Vision API." [Online]. Available: <https://cloud.google.com/vision/>. [Accessed: 20-Dec-2017].
- [43] "Google Cloud Vision Text Recognition." [Online]. Available: <https://cloud.google.com/vision/docs/ocr>. [Accessed: 21-Dec-2017].
- [44] "Google Developers." [Online]. Available: <https://developers.google.com/>. [Accessed: 25-Dec-2017].
- [45] F. van Diggelen and P. Enge, "The World's first GPS MOOC and Worldwide Laboratory using Smartphones " in *28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015)*, 2015, pp. 361 - 369.
- [46] "Longdo Map เพิ่มข้อมูลป้ายและสายรถเมล์ ขสมก. กรุงเทพฯ." [Online]. Available: <https://map.longdo.com/news/201802-longdomap-BMTA-buses-information>. [Accessed: 17-Mar-2018].
- [47] "แผนที่ Longdo Map แผนที่ประเทศไทย ใช้งานง่าย ละเอียด." [Online]. Available: <https://map.longdo.com/>. [Accessed: 17-Mar-2018].



ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



# An Automatic Bus Route Number Recognition

Pitchakorn Wongta, Thananop Kobchaisawat, and Thanarat H. Chalidabhongse  
 Computer Graphic and Computer Imaging Laboratory  
 Department of Computer Engineering, Faculty of Engineering,  
 Chulalongkorn University, Bangkok, Thailand.

E-mail: pitchakorn.w@student.chula.ac.th, thananop.k@student.chula.ac.th, thanarat.c@chula.ac.th

**Abstract**—This paper presents a bus route number recognition using image processing and machine learning techniques. The prototype is designed to extract bus route number from scene images and synthesize audio of bus number for people with low vision. The proposed system consists of two parts, text detection and digit recognition. Text detection based on MSERs is used to find and segment text area from input images. Digit recognizer based on Convolutional Neural Network is used to recognize digits on the detected text area. Finally, the recognized digit is converted to audio which can help the visually impaired people (low vision) to take the correct bus. The experimental results of the proposed method presented in precision, recall and F-measure in text detection and acquire 73.47% accuracy of a digit segment in self-collect Thai bus dataset.

**Keywords:** image processing; machine learning; bus route number recognition; visually impaired; low vision; convolutional neural network

## I. INTRODUCTION

According to World Health Organization (WHO) [1], there are 285 million visually impaired people that can be separated into 2 groups, 39 million are blind, and 246 million of whom have low vision. Blindness is complete or nearly complete vision loss, while low vision can see partially with a limited narrow angle. According to WHO, 'low vision' is defined as visual acuity of less than 6/18 but equal to or better than 3/60, or a corresponding visual field loss to less than 20°. The number of people with low vision has increased as a result of the aging population, and some deceases such as glaucoma and diabetic retinopathy which are increasingly important causes of low vision.

When their vision cannot be improved by using conventional eyeglasses, surgery, or medical treatment, people with low vision may find it difficult or impossible to accomplish activities such as reading, writing, shopping, watching television, recognizing faces, or even commuting. They need to learn how to make the most of their remaining sight and find new ways of doing daily activities. Low vision aids come in and range from simple hand-held lenses to electronic devices. Different ones are needed for different visual tasks. However, some aids are inconvenient to carry since it is too large while the other specifically designed portable aids are too expensive. These problems inspire us to develop a visual aid application on smart phone by using advanced mobile imaging technologies to enhance vision-related quality of life for people with low vision.

We conducted an interview with low vision people at the King Chulalongkorn Memorial Hospital's Low Vision Clinic. Many of them need care-takers to help in everyday activities, while some young aged people can take the benefit of smart phone to help them doing things independently such as reading sign by using mobile phone capturing picture and zoom in to read. However, one of the major difficulties for low vision people is to commute independently using public transportation. They can blurry see and recognize buses that are approaching the bus stop, but reading bus route number is still challenging to them. This leads to our motivation to develop an automatic bus route number reader which helps them recognize bus number from captured image like Fig. 1.

This paper presents a proposed bus route number recognition system. We use MSERs based text detection algorithm combined with MSERs segmentation to acquire each character. The extracted characters are then recognized by well-trained convolutional neural network (CNN). Finally, the recognize digits are then converted into synthetic sound.



Fig. 1. Example of an image of a bus approaching a bus stop captured by a conventional mobile phone.

The rest of this paper is organized as follows: In Section II, we present a survey of vision-base bus detection and bus route number recognition. Our proposed system overview is illustrated in Section III. Section IV describes text detection and MSER segmentation, as well as, the digit recognition using convolutional neural network. In Section V, the experimental setup and results are reported. Finally, conclusion and future work is described in Section VI.

## II. RELATED WORKS

In literature, there are a number of automatic bus number recognition systems; however, most of them use active sensors such as GPS tracking system, RFID, etc. For vision based approaches, Guida et al. [2] proposed the bus line number detection and recognition. Their method locates the main geometric entities of the bus facade through a cascade of classifiers based on Adaboost, which also used in face detection method proposed by Viola and Jones [3] and then refining the matching with robust geometric matching. The perspective correction technique is employed on the extracted digit area to produce perspective corrected digits. Then convert it to the HSV color space and applied thresholding operation to segment each digit. Finally, the OCR is used to recognize the digit and converted into voice notification for the visually impaired people.

Pan et al. [4] proposed another vision based method which detects information from the front of the bus by using the camera at bus stop. This system can detect the incoming bus route number and notify the visually impaired people in voice notification. Their system can be divided to 2 subsystems: bus detection and bus route number detection. In the bus detection, HOG and SVM are used to detect the bus location. In the bus route number detection, they detected bus route number and information from LED panel. They use adjacent character grouping and canny edge detection to find candidate regions. The Haar-like feature is extracted from edge distribution and stoke orientation maps. The acquired features are fed into Adaboost to classify each component. Finally, they use the Optical Character Recognition (OCR) software combined with text-to-speech synthesizer to produce an audio.

Tsai et al. [5] also proposed an automatic bus detection system to help visually impaired people to know bus route number. Their system including 3 subsystems: bus moving detection, bus panel detection, and text detection. In bus moving detection, they use Modify Adaptive Frame Differencing (MAFD) to extract the moving objects. If the difference between previous and current frame is small, current frame is skipped. Conversely, if the difference is moderate, moving object is close to the camera. In case of the moving object is big enough, an automatic threshold algorithm is applied on the different image to detect the moving objects which can be represent by a bounding box. If the interested bounding box is larger than predefined value, it classifies as a bus. Normally, the bus panel always located at the top of bus so, in bus panel detection part, the upper part of bounding box is selected as region of interest. Finally, they use background distribution and border thresholding to find the area of text which is converted to voice notification.

The three systems surveyed above deal with standardized bus with same color and same pattern both layout of the bus facade and line number format. Unfortunately, there are so many types of public buses in Bangkok with no standard format. There are so many variations in bus color, layout of the number plate, color and font of route number, appearance of route number, i.e., some are plain painted plates, some are painted plate under glass cover, some are digital LED panels, etc. Fig. 2 shows variations of Bangkok public buses. In addition, it is

known worldwide that Bangkok has some of the worst traffic jam. The scene images around bus stop area are typically quite complex. This creates occlusion; bus image region is typically incomplete. These are our challenges.



Fig 2. Samples of variations in Bangkok buses.

## III. SYSTEM OVERVIEW

Since the bus recognition and segmentation is still challenged as stated above. In this paper, we assume the bus route number panel already detected. Our system consists of two stages: text detection and digit recognition. The overview of the system is shown in Fig. 3. First, we enhance the input image by applying perspective transform to correct the geometric distortion. The output will be aligned rectangle of the route number panel. The MSERs based text detection is then used to localize the text location in the panel region. Then, MSERs based segmentation algorithm is employed to segment each character which is fed into our convolutional neural network digit recognizer to recognize each digit. Finally, the recognition output is converted into the audio.

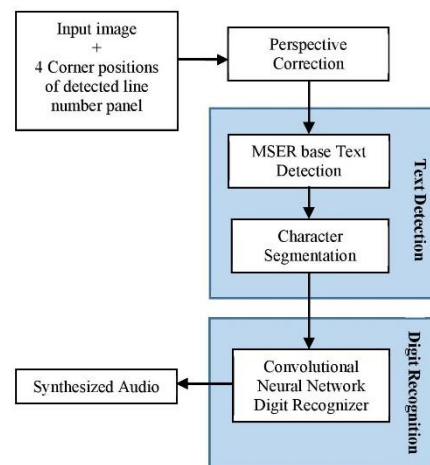


Fig. 3. The flowchart of proposed bus route number recognition system.



#### IV. PROPOSED METHOD

The proposed automatic bus route number recognition system consists of 3 main steps: image pre-processing, text detection and segmentation, and digit recognition and voice synthesis.

##### A. Pre-processing

Our pre-processing step begin by cropping the bus panel area from the provided four corner points. We then apply perspective correction on bus panel area to correct the skew as shown in Fig. 4. The output will be an aligned rectangle of the route number panel.

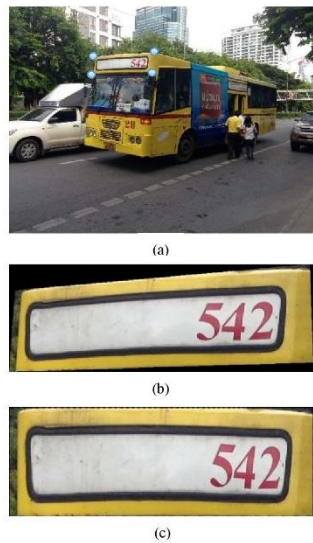


Fig. 4. (a) input image with specified 4 corner points of route number panel, (b) cropped panel image (c) perspective correction result.

##### B. Text Detection

In this step, the text in the bus panel region should be detected and segmented. The details are shown in text detection diagram in Fig. 5.

###### 1) Text Detection

In text detection part, we use OpenCV text detection [6]. This method is based on the algorithm proposed by Lukás Neumann & Jiri Matas [7] and Gomez L. and Karatzas D. [8]. This algorithm is based on MSERs and adopts trained classifier for character detection. The results are drawn in text character grouping. Our proposed method applies this algorithm to find a region of text on the bus panel image.

The results of text detection are presented in Fig. 6. The text areas are then cropped.

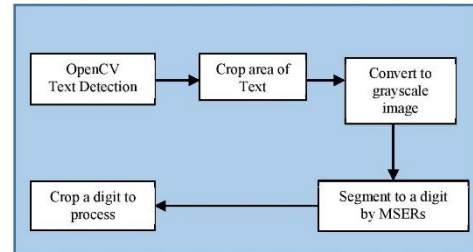


Fig. 5. The diagram of text detection.



Fig. 6. Example results of text detection.

###### 2) Digit Segmentation

We adopt Maximally Stable External Regions (MSERs) to segment characters from detected text area. First of all, we convert the output from text detection to grayscale and use MSERs to segment characters. The example of text segmented by MSERs are shown in Fig. 7. MSER is a technique to find stable region in image. It is popular in object recognition and stereo matching. This method was proposed by Matas et al. [9].

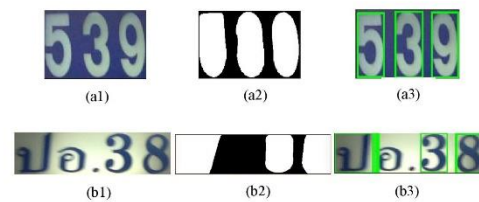


Fig. 7. Character segmentation using MSERs. (a1) and (b1) are input images, (a2) and (b2) results after applying MSERs, (a3) and (b3) the segmented results.

### C. Digit Recognition

After character segmentation, we then perform the recognition. Our convolutional neural network digit recognizer is trained by The Street View House Number (SVHN) dataset. All of sub-steps of digit recognition are shown in Fig. 8.

Convolutional neural network is a machine learning algorithm developed from neural networks. It is very famous classifier used in many applications such as object detection, video analysis, and natural language processing.

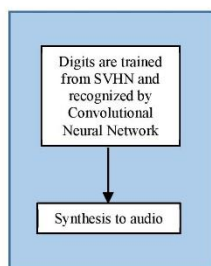


Fig. 8. The system diagram of digit recognition

In this paper, we apply convolutional neural network proposed in [10][11] to train digit dataset and recognized digit. The output has 11 classes, 0-9 and non-digit. The result is probability of a digit class.

## V. EXPERIMENTAL RESULT

In this section, we present the experimental results and discussion. First, we begin with dataset detail. Second, we provide result discussion.

The proposed method was implemented experimentally by Code::Blocks IDE on Ubuntu operating system on Intel® Core™ i7-4720HQ CPU @ 2.60GHz, NVIDIA Geforce GTX 960M.

### A. Bus Dataset

The dataset is collected by capturing 99 images of Thai buses in Bangkok area during the daytime. The images are captured when buses are approaching the bus stop. The dataset contains quite a variation of bus images, for examples, low contrast, reflection from bus panel cover glasses, blur images and various viewpoints.

### B. Results and Discussion

To demonstrate the performance of our proposed system, we conducted two experiments. The first experiment is text detection. After the pre-processing was done, we use text detection from OpenCV to find region of text. If region of text contains bus route number, we decide that result is correct. It is shown in Fig. 9. The evaluation of text detection is calculated in precision, recall, and F-measure which are presented in Table I.

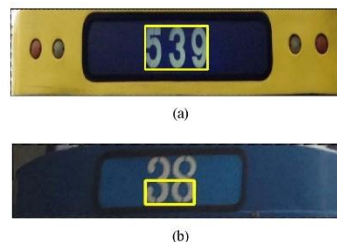


Fig. 9. Text detection results (a) is a correct one (b) is an incorrect one.

TABLE I. TEXT DETECTION RESULT

Text Detection		
Precision	Recall	F-Measure
0.78	0.67	0.73

The problem of text detection causes by variations of Bangkok bus panels due to differences in colors, fonts, and styles as shown in Fig. 10. Other kinds of variation of images that affect the detection are low contrast, reflection, and motion blur. These make finding region of bus route number difficult.



Fig.10. Variations of bus panels.

Second experiment is to evaluate digit recognition. Region of text from text detection is segmented using MSER. The results include both text and digits. We conduct an experiment by recognize only digits without Thai character in region of text. If the result has Thai character, it refers to non-text presented by nullity value as shown in Fig. 11.

The result of the second experiment is presented in confusion matrix shown in Table II. From the 66 extracted bus panel images by text detection, we adopt MSERs based segmentation method which acquired 98 segmented characters. It yields better results it clearly images and appropriate contrast than low contrast, reflective, and blur images.

TABLE II. CONFUSION MATRIX OF DIGIT RECOGNITION

		Predicted											
		1	2	3	4	5	6	7	8	9	0	Non-digit	
Segmented Character	1	1	-	-	-	-	-	-	-	-	-	-	2
	2	-	8	-	-	-	-	-	-	-	-	-	2
	3	-	-	5	-	-	-	-	-	-	-	-	1
	4	-	-	-	8	-	-	-	-	-	-	-	2
	5	-	-	-	-	12	-	-	-	-	-	-	-
	6	-	-	-	-	-	5	-	-	-	-	-	2
	7	-	-	-	-	-	-	1	-	-	-	-	-
	8	-	-	-	-	-	-	-	1	-	-	-	4
	9	-	1	-	-	-	-	-	-	4	-	-	-
	0	-	-	-	-	-	-	-	-	-	-	-	2
	Non-digit	9	-	-	-	1	-	-	-	-	-	-	27

From the 98 segmented characters, our algorithm gives 72 correctly detected. The digit recognition accuracy is 73.47 percent as presented in Table III.

TABLE III. DIGIT RECOGNITION ACCURACY

Digit Recognition		
NO. of images segmented	Correctly Detected	Accuracy
98	72	73.47%

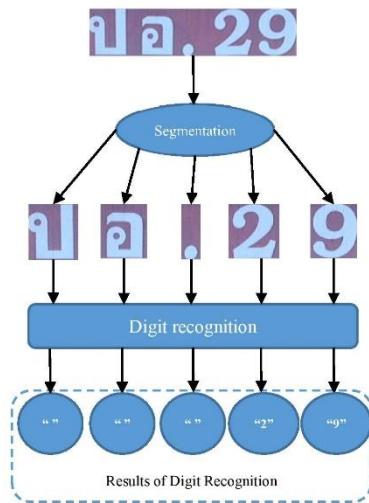


Fig. 11. Our digit recognizer will recognize only digits. Other characters will be classified into an unknown class ("??").

We found a problem with some Thai characters which are similar to digits. Thus, the recognizer yields the wrong results as shown in Fig. 12.

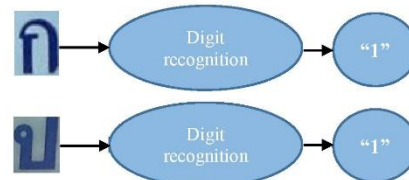


Fig. 12. Some example of Thai characters that are similar to digits.

VI. CONCLUSION AND FUTURE WORK

In this paper, we present an automatic bus route number recognition. The project aims to develop a visual aid for people with low vision in reading bus line number. We hope with this kind of application will be helpful to them in commuting using public transportation by their own and reduce chances of accidents when they want to take a bus. From our experiments,

we can achieve 0.78 of precision, 0.67 of recall and 0.73 of F-measure in text segmentation. Second, we recognize only digits without Thai character and can achieve 73.47% in cropped bus panel.

For the future work, we want to modify text detection to improve an accuracy of the digit recognition. We also plan to eventually implementing the algorithm on smartphone platform to help people with low vision.

#### REFERENCES

- [1] "Visual impairment and blindness." [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs282/en/>. [Accessed: 15-February-2014].
- [2] C. Guida, D. Comanducci, and C. Colombo, "Automatic Bus Line Number Localization and Recognition on Mobile Phones—A Computer Vision Aid for the Visually Impaired," in *Image Analysis and Processing – ICAP 2011*, G. Maino and G. L. Foresti, Eds. Springer Berlin Heidelberg, 2011, pp. 323–332.
- [3] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
- [4] H. Pan, C. Yi, and Y. Tian, "A primary travelling assistant system of bus detection and recognition for visually impaired people," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–6.
- [5] C. M. Tsai and Z. M. Yeh, "Text detection in bus panel for visually impaired people "seeing" bus route number," in *2013 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2013, vol. 03, pp. 1234–1239.
- [6] "Scene Text Detection." [Online]. Available: <http://docs.opencv.org/3.0-beta/modules/text/doc/erfilter.html#erfilter>.
- [7] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3538–3545.
- [8] L. Gómez and D. Karatzas, "Multi-script Text Extraction from Natural Scenes," in *2013 12th International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 467–471.
- [9] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.
- [10] T. Kobchaisawat and T. H. Chalidabhongse, "Thai text localization in natural scene images using Convolutional Neural Network," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, pp. 1–7.
- [11] T. Kobchaisawat and T. H. Chalidabhongse, "A method for multi-oriented Thai text localization in natural scene images using Convolutional Neural Network," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2015, pp. 220–225.

# Vision-Based Bus Route Number Reader for Visually Impaired Travelers

Pitchakorn Wongta and Thanarat H. Chalidabhongse  
Department of Computer Engineering, Faculty of Engineering,  
Chulalongkorn University, Bangkok, Thailand.  
E-mail: pitchakorn.w@student.chula.ac.th, thanarat.c@chula.ac.th

**Abstract**—In this paper, we present vision-based bus route number reader by using machine learning and image processing techniques. This prototype system is designed for people with low vision to recognize bus route number from natural scene images. Our proposed system consists of three stages: bus panel detection, image enhancement, and bus number recognition. The bus panel detection is done using faster RCNN. The pre-processing is applied to enhance bus panel images. Next, the bus route number is recognized using Google Cloud Vision API. Finally, the recognized bus route number is translated to synthesized audio. The prototype system can be applied to develop a mobile application or embedded system at the bus stop to aid people with low vision when taking the public bus transportation. Experimental results presented in precision, recall and f-measure in bus panel detection stage and obtain 67.9 percent accuracy of bus number recognition.

**Keywords:** bus route number recognition; visually impaired; low vision; assistive technology, faster RCNN; tensorflow object detection

## I. INTRODUCTION

According to the International Classification of Diseases, vision function is classified into 4 broad categories: normal vision, moderate vision impairment, severe vision impairment, and blindness. “Low vision” is a term that groups both moderate and severe vision impairments. People with low vision can partially see. Causes of low vision include hereditary conditions, eye injuries, eye diseases such as glaucoma, diabetic retinopathy, and especially age-related macular degeneration. Low vision can occur at any age, but by far the greatest number of people who are partially sighted are the elderly. The Low Vision Association (Thailand) estimated in 2013 there were about 300,000 people in Thailand who have low vision. According to the World Health Organization (WHO)’s fact sheet updated in 2017 [1], there are approximately 217 million people having moderate to severe vision impairment and could triple by 2050 due to population growth and ageing.

Imagine what life would be like if we lost a certain amount of eyesight and could not be corrected with eyeglasses, contact lenses, or laser surgery. With low vision, it is hard and suffering to do many of normal tasks such as reading, shopping, and travelling. People with low vision need to use assistive devices which enable them to use their remaining vision more effectively and do things they thought were no longer doable.

These devices include hand-held magnifiers, stand electronic magnifiers that can adjust images and colors, computerized text-to-speech devices, etc. These devices are used for close-up, detail viewing suitable for near-sighted tasks such as reading, writing, and crafting. To see objects in the distance for activities that are farther away such as watching television, seeing movies or sport events, reading street signs, or identifying train or bus route numbers, common devices in today’s market are hand-held and spectacle-mounted telescopes. However, these telescopes often provide a limited field of view and quite dark illumination, which can be frustrating. They also require good motor control to focus the lenses, and to be steady to keep it in focus.

It has become accepted in today’s world that cameras and smartphones are everywhere. This has inspired us to develop a vision-based recognition system that can assist the low vision people to maintain their independence for activities of daily living such as travelling using the public transportation system. In Bangkok, while the rail systems recently have been developed and being expanded, the huge network of public buses serviced by the Bangkok Mass Transit Authority (BMTA) and privately-owned companies provides the incredibly cheap and easily accessible way to travel in the city. Normally, at the bus stop, the people with low vision need to ask other people of what route number the bus is approaching. Without any assistance, they may get lost or even get injured by an accident.

By leveraging the highly developed computer vision and machine learning technologies, in our previous work, an automatic bus route number reading systems were proposed [2] to aid low vision people to able to travel using public buses more safely and confidently. The prototype we developed as an assistive system to help the people with low vision to recognize the bus route number from a captured image, see Fig. 1. The system began with some pre-processing and manually cropped the bus panel region. Then, a text-detection based on the Maximally Stable Extremal Regions (MSERs) [3], which is a technique to find stable regions in the image, was used to locate the bus route number area and segment it into regions of characters. The CNN number recognizer, which was trained by the Street View House Number (SVHN) [4] [5] dataset, was applied to recognize the route number. Lastly, the voice of route number was synthesized. However, due to variations of image viewpoints, illuminations, brightness, contrast, and noises, the recognition rate of the previous system should be improved.

This paper presents an improvement of the system by adding an automatic bus panel detection as well as proposes different techniques for route number recognition.

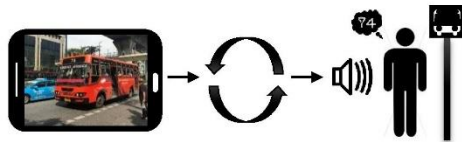


Fig. 1. The bus route number reader processes captured bus image taken by mobile phones and synthesizes voice to aid people with low vision at bus stops.

The rest of the paper is organized as follows: Section II provides overview of the related works. Section III introduces our system overview. The details of proposed techniques in bus panel detection and route number recognition are described in Section IV. Section V presents the experiments and Section VI concludes the paper.

## II. RELATED WORKS

The existing approaches for bus route number detection and recognition can be categorized into two groups: active sensor-based and vision-based methods. The active sensor-based methods utilize wireless sensors such as Global Positioning Systems (GPS) and Radio Frequency Identification (RFID). Noor et al. [6] proposed a bus detection device for the blind using RFID. Their system utilizes RFID to communicate between buses and bus stops. When the bus approaches the bus stop, the bus information can be retrieved. Kalbani et al. [7] also proposed bus detection system for blind people using RFID. Their system can be divided to two subsystems: station and bus subsystems which are connected to a database. When the bus approaches the station, the station subsystem reads bus information and converts to voice for blind passengers. In addition, the number of the blind passengers as well as their destination information, which are registered at the station earlier, will be sent to the bus.

For vision-based methods, Guida et al. [8] proposed an automatic bus line number localization and recognition on mobile phones. Adaboost is used as a bus façade classifier. The perspective correction technique is applied on the bus façade to correct the geometric distortion. The predefined bus template is applied to extract a bus number. After all those steps, color segmentation in HSV color space and thresholding are applied to segment the bus numbers into individual characters. The Optical Character Recognition (OCR) is applied to recognize the bus number. Finally, the bus numbers are synthesized to voice for visually impaired people.

A primary travelling assistant system of bus detection and recognition for visually impaired people was proposed by Pan et

al. [9] This system detected the approaching bus and warned the visually impaired people by using sound. Their proposed system can be separated into two subsystems: the bus detection and the bus number detection subsystems. Histogram of the Oriented Gradient (HOG) and Support Vector Machine (SVM) were applied to detect the bus location in the bus detection system. The detected bus region was fed to the bus number recognition system, which could be separated to coarse and fine stages. In the coarse stage, adjacent character grouping and canny edge detection were used to filter out non-number regions. Then, the rest of regions was forwarded to the fine stage classifier based on the Adaboost to acquire the bus number region. Finally, OCR software and speech synthesis were used as transcribing the bus number images into speech.

Lee et al. [10] proposed automatic number recognition for bus route information aid for the visually-impaired. This system can be divided to two subsystems: the bus number detection and the bus number recognition subsystems. In the bus number detection subsystem, they adopted the census transform feature and fed result to the Adaboost, which was trained by the number from license plate number dataset. In the bus number recognition subsystem, the thresholding was applied to segment the number to be individual component. Then, the gradient feature was extracted from each component and forwarded to the multiclass SVM. The multiclass SVM was used as the number class classifier. Lastly, they used the text to speech from Google to convert text to speech.

The bus number detection via motion and YCbCr features was proposed by Cheng et al. [11]. This system began with the bus area detection. Adaptive frame differencing was applied to find the bus area from the moving bus. Aspect ratio correction properly adjusted the aspect ratio of bus panel. Finally, YCbCr color space is used to find area of bus number. The other researches for bus number detection using YCbCr features were proposed by Tsai et al [12]. This system applied YCbCr color space to find the bus number from moving bus. Connected component labeling was applied to find the bus number candidates. The bus number was represented as bounding-boxes. However, many noises were detected. To remove noises from irrelevant bounding-boxes, localization method was applied. Finally, the bus numbers were detected.

In summary, the active sensor systems provide higher accuracy, but limits in distance. To increase the distance, expensive high-performance equipment is required.

In vision-based systems, to the best of our knowledge, the previous systems reviewed above work based on assumptions of standard templates in term of bus body, panel, and route number format, pattern, and color. Apparently, due to a huge variation in Bangkok's buses, the existing vision-based techniques will not work. The variations of buses in Bangkok include variations in body types, size, and color of the buses, variations in position, size, color, and material of the route number panels, as well as variations in font, size, color, and types of the route numbers. Fig. 2 shows the examples of the bus variations in Bangkok. This paper proposed learning-base approach in detecting bus and bus panel by training with self-collected bus dataset.



Fig 2. The example of the bus variations in Bangkok buses.

### III. SYSTEM OVERVIEW

Our system overview shows in Fig. 3. It consists of three main stages: the bus panel detection, the image enhancement, and the bus number recognition. In the first stage, the image regions of bus panels are localized from scene images like ones shown in Fig. 2 using Faster RCNN. Then, a sequence of image processing technique is used to enhance the bus panel image to suite the next step. Lastly, the OCR from Google Cloud Vision API is applied to recognize the bus route number from the enhanced panel image region before the synthesized sound of the bus route number is done.

### IV. PROPOSED METHOD

#### A. Bus Panel Detection

Bus panel detection is trained using TensorFlow object detection [13], which uses pre-trained-weight ResNet101 based on faster RCNN model. The COCO dataset [14] was trained to be pre-trained-weight model. This trained model is used for testing the detection in our bus panel dataset. The results from the bus panel detection are shown in Fig. 4. The bus panel is detected into the form of bounding box.

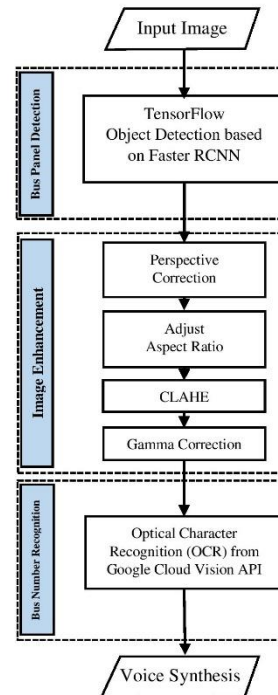


Fig. 3. The system overview of proposed system.

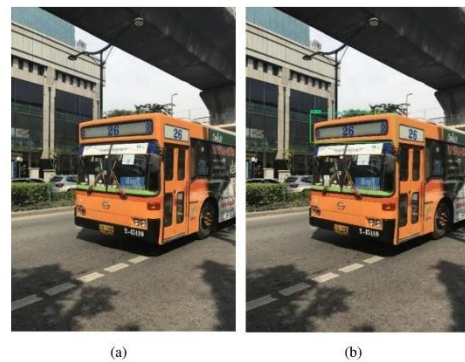


Fig. 4. The example result of bus panel detection. (a) input image, (b) Output image, which the bus panel area is detected, represented in a green rectangle.

Faster RCNN [15] is machine learning algorithm, which was developed from Fast RCNN combined with Region Proposal Network (RPN). It is powerful classifier for object detection because it produces higher accurate result and it is faster than fast RCNN. ResNet101 is the pre-trained network, which processed the good detecting results in ImageNet and COCO datasets [16], and it was applied to the Faster RCNN.

### B. Image Enhancement

Possible area of bus number, which is green rectangle, is automatically cropped from the bus panel detection. Next, perspective correction is applied to recover the perspective distortion. The aspect ratio correction is also applied. Fig. 5. shows output image of this processing.

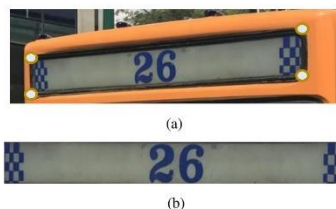


Fig. 5. (a) Four corners of bus panel are detected from the bus panel image obtained from the Faster RCNN (b) Bus panel after perspective and aspect ratio correction.

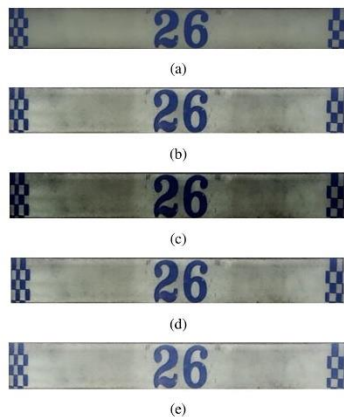


Fig. 6. Illustration of image enhancement steps (a) input image, (b) pre-processing with CLAHE, (c) pre-processing with CLAHE + Gamma Correction ( $\gamma = 0.5$ ), (d) pre-processing with CLAHE + Gamma Correction ( $\gamma = 1.0$ ) and (e) pre-processing with CLAHE + Gamma Correction ( $\gamma = 1.5$ )

The Contrast Limited Adaptive Histogram Equalization method (CLAHE) [17] is an enhancement technique, which is adjusts image contrast for improving image details. Then, we

find mean value of Y in the 8-bit YCbCr color spaces from the images. In gamma correction [18], which known as the power law transform, we use the mean value of Y to adjust brightness from gamma values. The example of image enhancement using CLAHE and gamma correction with varied gamma values are shown in Fig. 6.

### C. Bus Number Recognition

The Google Cloud Vision API [19] is widely used in the image processing and the computer vision fields. It enables developers to receive the content of an image by its machine learning models. It has many features to analyze an image such as label detection, face detection, logo detection, OCR, etc.

In the bus number recognition system, we use the OCR from the Google Cloud Vision API which can recognize texts in many languages. This OCR will upload the image to the Google Cloud. Their translation and vision API extract the bus number text from the output image of the image enhancement step.

## V. EXPERIMENTAL RESULT

In this section, we present the experimental results and discussion. First, we start discussing with the dataset details. Second, we explain our results and discussion.

The proposed system was implemented on Ubuntu operating system on Intel® Core™ i7-4770 CPU @ 3.40GHz, NVIDIA Geforce GTX 1080 Ti.

### A. Dataset

Dataset images were collected by capturing bus image using mobile phones. The current dataset contains 540 images taken in day time while buses were approaching the bus stops in Bangkok metro area. The dataset contains wide variations in clarity, contrast, illumination, viewpoint, and scene complexity. Some images may contain more than one bus. The dataset is divided into two sets: one for training and the other one for testing. The training set consists of 432 randomly selected images used for training the bus panel classifier. The testing set is the remaining 108 images used for evaluating the panel classifier.

### B. Results and Discussion

To evaluate our system performance, we conducted the experiment into two parts: the bus panel detection and bus number recognition experiments. First, we fed the input image to our bus panel detection system. The bus panel regions will be detected by our system. The results were evaluated by the Intersection over Union (IoU) [20] that is a comparison between prediction and ground truth areas.

The IoU is widely used for evaluating the accuracy of object detection. The famous competition like PASCAL VOC competition also uses the IoU. The equation (1) is formula for



calculating the IoU. In our system, the area of overlap calculated from intersecting area between the prediction and the ground truth and the area of union calculated from union area between the prediction and the ground truth.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

The example of the IoU score is shown in Fig. 7. Our scoring condition considers the correct prediction when the IoU has score more than 0.7.

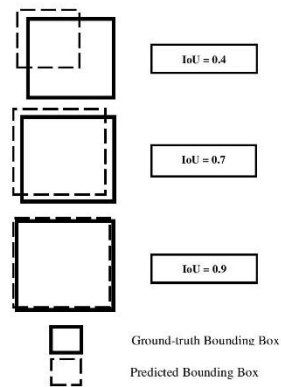


Fig. 7. An example of intersection over union scoring for various bounding boxes.

The first experimental result shows that 112 bus panels correctly detected. The bus panel detection gives 114 prediction boxes from 108 evaluation bus images. The amount of ground truth has 113 images. The errors can be divided to two cases: incorrectly detected prediction and missed prediction. Our system incorrectly detected 2 prediction boxes and miss 1 bus panel. The result is presented in precision, recall and F-measure. It is shown in Table I.

TABLE I. BUS PANEL DETECTION RESULT

Bus Panel Detection		
Precision	Recall	F-Measure
0.982	0.991	0.986

The second experiment was evaluated the bus number recognition. After the bus panel detection, the segmented bus panel image region is proceeded to the image enhancement stage before entering the bus number recognition stage. This is done using the OCR from the Google Cloud Vision API. We scoped the experimental evaluation without Thai character and focused on only the number recognition.

We experimented by using various pre-processing techniques and parameters before doing the OCR. The recognizing results are shown in Table II.

The 112 bus panels, which were correctly detected from the bus panel detection stage, were processed in the perspective and aspect ratio correction steps. Then, we apply the 112 bus panels with eight pre-processing techniques. First technique used ordinary bus panel images to evaluate the accurate recognition. Second technique enhanced the bus panel images with the gamma correction. Third technique used the CLAHE method to modify the bus panel images. Fourth technique used the gamma threshold with the mean value of Y from YCbCr color spaces to adjust the bus panel images. Fifth technique used the CLAHE and the gamma correction to enhance the bus panel images. Sixth, Seventh and Eighth techniques used the gamma threshold and the CLAHE to enhance the bus panel images. The best result is the 8<sup>th</sup> technique pre-processing. It gave 67.9 percent of the accurate recognition.

TABLE II. BUS NUMBER RECOGNITION RESULT

Bus Number Recognition						
	Gamma Correction			CLAHE	Correctness	Accuracy
	$\gamma = 1.5$	$\gamma = 1.0$	$\gamma = 0.5$			
1	-	-	-	-	71	63.4%
2	✓	-	-	-	69	61.6%
3	-	-	-	✓	64	57.1%
4	$\mu_Y < 180$	$\mu_Y = [180, 220]$	$\mu_Y > 220$	-	67	59.8%
5	✓	-	-	✓	71	63.4%
6	$\mu_Y < 100$	$\mu_Y = [180, 200]$	$\mu_Y > 200$	✓	66	58.9%
7	$\mu_Y \leq 100$	$\mu_Y > 100$	-	✓	72	64.2%
8	$\mu_Y \leq 180$	$\mu_Y > 180$	-	✓	76	67.9%



Fig. 8. An example of variation of bus panel such as single character, reflection, low-light image and more font and panel styles.

From our experimental result, we discovered 12 of 112 images that our method could not recognize the bus numbers. There are various reasons of the errors such as font styles, noises, incomplete characters, reflections, blur and low-light image. The other main reason is that most of text detection and recognition processes assume any text or word has connected characters as a chunk. Therefore, a single bus number cannot be detected or recognized because of that reason. The examples of various error are shown in Fig. 8.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed the vision-based bus route number reader to aid people with low vision while taking the public bus transportation. We wish them to be able to travel on their own more safely and confidently. The bus panel detection using Faster RCNN achieved 0.982 precision, 0.991 recall and 0.986 F-measure tested on 108 test images which contain 113 bus panels. After obtain bus panel, the OCR is done after some image enhancement. The best recognition rate we get at presents is 67.9 %.

For the future work, we need to improve the recognition by incorporating some prior knowledge of geographic information such as bus routes and user location as well as working on better model of the digit recognizer.

## REFERENCES

- [1] "Visual impairment and blindness." [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs282/en/>. [Accessed: 21-Jan-2018].
- [2] P. Wongta, T. Kobchaisawat, and T. H. Chalidabhongse, "An automatic bus route number recognition," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2016, pp. 1–6.
- [3] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.
- [4] T. Kobchaisawat and T. H. Chalidabhongse, "Thai text localization in natural scene images using Convolutional Neural Network," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, pp. 1–7.
- [5] T. Kobchaisawat and T. H. Chalidabhongse, "A method for multi-oriented Thai text localization in natural scene images using Convolutional Neural Network," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2015, pp. 220–225.
- [6] M. Z. H. Noor, I. Ismail, and M. F. Saaid, "Bus detection device for the blind using RFID application," in *5th International Colloquium on Signal Processing Its Applications, 2009. CSPA 2009*, 2009, pp. 247–249.
- [7] J. A. Kalbani, R. B. Suwailam, A. A. Yafai, D. A. Abri, and M. Awadalla, "Bus detection system for blind people using RFID," in *GCC Conference and Exhibition (GCCCE), 2015 IEEE 8th*, 2015, pp. 1–6.
- [8] C. Guida, D. Comanducci, and C. Colombo, "Automatic Bus Line Number Localization and Recognition on Mobile Phones—A Computer Vision Aid for the Visually Impaired," in *Image Analysis and Processing – ICIAP 2011*, G. Maimo and G. L. Foresti, Eds. Springer Berlin Heidelberg, 2011, pp. 323–332.
- [9] H. Pan, C. Yi, and Y. Tian, "A primary travelling assistant system of bus detection and recognition for visually impaired people," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–6.
- [10] D. Lee, H. Yoon, C. Park, J. Kim, and C. H. Park, "Automatic number recognition for bus route information aid for the visually-impaired," in *2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 2013, pp. 280–284.
- [11] C. C. Cheng, C. M. Tsai, and Z. M. Yeh, "Detection of Bus Route Number via Motion and YCbCr Features," in *2014 International Symposium on Computer, Consumer and Control*, 2014, pp. 31–34.
- [12] C.-M. Tsai and Z.-M. Yeh, "Detection of Bus Routes Number in Bus Panel via Learning Approach," in *Intelligent Information and Database Systems*, N. T. Nguyen, B. Attachoo, B. Trawiński, and K. Somboonviwat, Eds. Springer International Publishing, 2014, pp. 302–311.
- [13] "Tensorflow Object Detection API." [Online]. Available: [https://github.com/tensorflow/models/tree/master/research/object\\_detection](https://github.com/tensorflow/models/tree/master/research/object_detection).
- [14] "Tensorflow detection\_model\_zoo." [Online]. Available: [https://github.com/tensorflow/models/blob/master/research/object\\_detection/g3doc/detection\\_model\\_zoo.md](https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md).
- [15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 91–99.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [17] "CLAHE (Contrast Limited Adaptive Histogram Equalization)." [Online]. Available: [https://docs.opencv.org/3.1.0/d5/daft/tutorial\\_py\\_histogram\\_equalization.html](https://docs.opencv.org/3.1.0/d5/daft/tutorial_py_histogram_equalization.html).
- [18] "OpenCV Gamma Correction." [Online]. Available: <https://www.pyimagesearch.com/2015/10/05/opencv-gamma-correction/>.
- [19] "Cloud Vision API." [Online]. Available: <https://cloud.google.com/vision/>.
- [20] "Intersection over Union (IoU) for object detection." [Online]. Available: <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>.

## ประวัติผู้เขียนวิทยานิพนธ์

นาย พิชชากร วงศ์ดีตะ เกิดเมื่อวันที่ 12 มีนาคม พ.ศ. 2535 ที่จังหวัดเชียงราย สำเร็จ การศึกษาหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ สำนักวิชา เทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่ฟ้าหลวง ในปีการศึกษา 2556 และเข้าศึกษาต่อในหลักสูตร วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะ วิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2557

